Thailand Statistician
January 2009; 7(1) : 13-27
http://statassoc.or.th
Contributed paper

Simple Latin Cubic Sampling +1 and -k Sampling Designs

Kamon Budsaba *[a], John J. Borkowski [b], and Kanlaya Boonlha [a]

[a] Department of Mathematics and Statistics, Thammasat University, Rangsit Center, Bangkok 12121, Thailand.

[b] Department of Mathematical Sciences, Montana State University, Bozeman, MT 59717 USA

***Author for correspondence**; e-mail: k_budsaba@hotmail.com

Received: 29 August 2008

Accepted: 20 November 2008.

Abstract

In this study we presented a new class of probability sampling designs, simple latin cubic sampling +1 sampling designs that were developed from simple latin square sampling designs by focus on three-dimensional, with the specific goals of deriving an estimator of the population total, true variances of these estimators, and estimators of these variances. And the Horvitz-Thompson estimation method will be the primary method used to generate these estimator. These designs when compared with simple random, stratified, and systematic sampling will provide estimator with smaller variance for simulation population with spatial correlation and assume that the survey region can be partitioned into three-dimensional grid of d^3 equalized three-dimensional.

Keywords: Horvitz–Thompson estimators, Latin square sampling, sampling designs.

1. Introduction

In many biological, sociological, agricultural and geological studies, the population of interest is a region that is partitioned into quadrates which represent the sampling units. The variable of interest is some characteristic that can be measured in each quadrate. After taking a sample of these quadrates by a particular sampling design and recording the measurement for the variable of interest, this information can be used to estimate a parameter of the population. Furthermore, suppose that the variable of interest has a positive spatial correlation which means neighboring units tend to have more similar values than units that are far apart. This type of spatial correlation is presented in many biological and geological populations. Therefore, it is desirable for present sampling units to be scattered throughout the population to assure a representative sample with good spatial coverage.

Sampling designs which produce an appropriate coverage of the population will increase the precision of the parameter estimator [1], and a spatially well-distributed sample may also be advantageous for estimating a population parameter. For example, if spatial correlation is presented, an estimator of abundance could be more precise when based on data arising from sampled quadrates that are well distributed over the study region. Unfortunately, simple random sampling of quadrates does not guarantee that a particular sample will provide good coverage over the region. In a one-dimensional population, systematic and stratified sampling with a random sample taken in each stratum is traditional sampling designs that cover the population region [2].

For two-dimensional populations new probability sampling designs that yield spatially well-distributed samples will be presented. These designs, called “simple latin square sampling ± k designs, were developed by Borkowski [3] from simple latin square designs. In these sampling designs the sample size equals the square root of the population size. These sampling designs which are alternatives to other sampling designs for two-dimensional populations can provide better sample coverage than either systematic or simple random sampling. Also, the corresponding estimators are generally more efficient than those obtained from the latter designs when the population is spatially correlated.

Suppose that a population is partitioned into a square grid of d^2 substrata according to two stratification criteria such that d strata are formed from each criterion. Historically, the term latin square sample refers to a sample of d substrata such that exactly one substratum from each of the square’s rows and columns is selected. A sampling unit is then randomly selected within each chosen substratum. This sampling design has been called a variety of names, including lattice sampling [4,5] latin square stratification [6] and multiple stratification [7]

Typically, the data may be collected from a one-dimensional or two-dimensional study region. Although previous research has been limited to one or two-dimensional problems, applications of three-dimensional volumetric data are becoming increasingly available in a wide range of scientific and technical disciplines. The spatial objects of our world have an intrinsic three-dimensional nature, and they have often been neglected in spatial systems and processes. We can expect such data to yield valuable insights about many important systems in our three-dimensional world. Accounting for uncertainty in three-dimensional shapes is important in a large number of scientific and engineering areas, such as biometrics, biomedical imaging, and data mining. For example, Geographical Information Systems (GIS), which map geometric data, are nowadays restricted to the handling of two-dimensional information. But increasingly the third dimension becomes more and more relevant for application domains like pollution control, water supply, soil engineering, mining, urban planning and aviation [8]. However, the statistics of these three-dimensional models have not been widely explored. The statistics are similar to those commonly found in other fields of the physical, biological, and earth sciences.

For these various reasons, we are interested in the extension of simple latin square sampling design principles and theory to applications in three-dimensional sampling frames. The research will focus on three-dimensional simple latin square sampling +1 designs with the specific goals of deriving an estimator of the population total, an estimator of the population mean, the true variances of these two estimators, and estimators of these variances. We will call the new sampling designs that "simple latin cubic sampling designs "denoted SLCS are modifications of sample latin square sampling +1 designs (SLSS+1) ,which were introduced by Munholland and Borkowski [9]. And the Horvitz-Thompson estimation method will be the primary method used to generate these estimators.

2. Simple Latin Cubic Sampling Designs

Simple latin cubic sampling designs (denoted SLCS) are modifications of sample latin square sampling designs (SLSS), which were introduced by Munholland and Borkowski [9]. The SLSS designs will be extended to three dimensions by applying simple latin square sampling design principles and theory to three-dimensional sampling frames, and the researcher's goal is to estimate the population total (τ). Assume that the survey region can be partitioned into a $d \times d \times d$ three-dimensional grid of d^3 equalized three-dimensional rectangular solid which will be referred to as "cubic". Thus the cubic are form the population of sampling units.

For each cubic, let *y* be the response of interest. The values of the population units with respect to the characteristic *y* under study will be denoted by y_1, y_2,....,y_N where $N= d^3$. Here y_i denotes the value of the unit bearing labels *i* with respect to the variable y. Suppose a sample of $n = d$ cubic are selected from the grid of d^3 cubics via a sampling design, and the y-values are recorded. Let the three dimensions be referred to as "layer", "row", and "column", and let (l, r, c) denote the cubic at layer = l, row = r, and column = c. The following algorithm contains the three steps for selecting the units when the sampling design is a SLCS.

Step 1: Take a SLSS of size d in the two-dimensional square of $d \times d$ quadrats. That is, generate a permutation of the numbers 1, 2 ,…, d and then sequentially assign the numbers in the permutation to the columns while moving down the rows.

Step 2: Generate a random permutation ($a_1, a_2, \ldots, a_d$) of the integers 1,2,…d.

Step 3: Assign the SLSS unit in row r to layer a_r for r = 1,2,…,d.

An example of selecting a SLCS when d = 4 will now be presented. In Step 1, a SLSS of units (r, c) = (1, 1), (2, 2), (3, 4), and (4, 3) was selected. Figure 1 shows the results for this SLSS with the •'s represent the sampling units.

	1	2	3	4
1	•			
2		•		
3				•
4			•	

Figure 1. An example of selecting units in SLSS.

A random permutation (3, 1, 2, 4) of the integers 1, 2, 3, 4 is generated. Layers 3, 2, 1, and 4 are then assigned to the SLSS units in rows 1 to 4, respectively. These layers yield a SLCS of (3, 1, 1), (1, 2, 2), (2, 3, 4), and (4, 4, 3) shown in Figure 2.

Layer 1

	1	2	3	4
1				
2		•		
3				
4				

Layer 2

	1	2	3	4
1				
2				
3				•
4				

Layer 3

	1	2	3	4
1	•			
2				
3				
4				

Layer 4

	1	2	3	4
1				
2				
3				
4			•	

Figure 2. Generating a SLCS from the SLSS in Figure 1.

3. Simple Latin Cubic Sample +1 Designs

A more general design can be easily defined by initially selecting an SLCS and subsequently drawing *k* additional units to give an SLSS+k where $1 \leq k \leq (d-1)$. An approach to increasing sample size, we focus on the simplest case with *k*=1. A SLCS+1 fall into the classical designs framework in that samples generated by the design have corresponding selection probabilities that are independent of *y*-values.

In these sampling designs, the sample size can be *d* +1 which is a modification of simple latin cubic designs. A simple latin cubic sample +1 design (SLCS+1) is composed of two sets of sampling units:

(i) *d* sampling units from a SLCS, and

(ii) one additional sampling unit which is a cubic selected from the d^3 - *d* remaining units.

Suppose a sample of $n = d + 1$ units are selected from the d^3 cubic via a sampling design, and for each cubic, let *y* be the response of interest. The following algorithm contains the steps for selecting the units in a SLCS+1 :

Step 1: Take a SLSS in the two-dimensional square of $d \times d$ quadrats.

Step 2: Generate a SCLS from this SLSS.

Step 3: Randomly select a cubic from the d^3- *d* remaining cubic.

An example of a simple latin cubic sample of size *d*=4 from a population of 64 units generated by the permutation (1, 2, 4, 3) yields the SLSS indicated with •'s in Figure 1. Let (3, 1, 2, 4) be a random permutation of the integers 1,2,3,4. Assign the SLSS unit in row 1 to layer 3, in row 2 to layer 1, in row 3 to layer 2, and in row 4 to layer 4 This generates a SLCS with units (*l , r, c*) = (3, 1, 1), (1, 2, 2), (2, 3, 4), and (4, 4, 3). The SLCS show in Figure 2. The additional unit (+) is randomly chosen from the remaining 60 units to give a simple latin cubic sample +1. The SLCS+1 design for *d* = 4 and the additional unit is shown in Figure 3.

Layer 1

	1	2	3	4
1				
2		•		
3				
4				

Layer 2

	1	2	3	4
1				
2				
3				•
4				

Layer 3

	1	2	3	4
1	•			
2				+1
3				
4				

Layer 4

	1	2	3	4
1				
2				
3				
4			•	

Figure 3. Example of SLCS+1 where d=4.

There are three major benefits to selecting the +1units with this algorithm:

(i) the +1 unit is an extension of the SLCS. That is, the additional one unit spans another layer, row, and column, and thereby continues to provide a sample of cubic that is spatially well-distributed,

(ii) the joint inclusion probability in the Horvitz-Thompson estimator is greater than 0 for all pairs of cubic. This guarantees that the estimator of the variance will be unbiased.

4. Simple Latin Cubic Sample - k Design

If suppose the researcher does not have enough resources to sample at least d + 1 units. Assume a maximum of d - k (k = 0, 1,.., d – 2) units could be sampled. Another class of designs that allows these smaller sample sizes are known as simple latin cubic sampling - k designs (SLCS-k). The units in a SLSS-k are a subset of sampling units from any SLSS. A SLCS-k can be generated by randomly selecting (d - k) points from a randomly generated SLCS.

An example of a simple latin cubic sample of size d = 4 with units (l, r, c) = (3, 1, 1), (1, 2, 2), (2, 3, 4), and (4, 4, 3) is shown in Figure 2. An example of a SLCS - 1 design for d = 4 is shown in Figure 4.

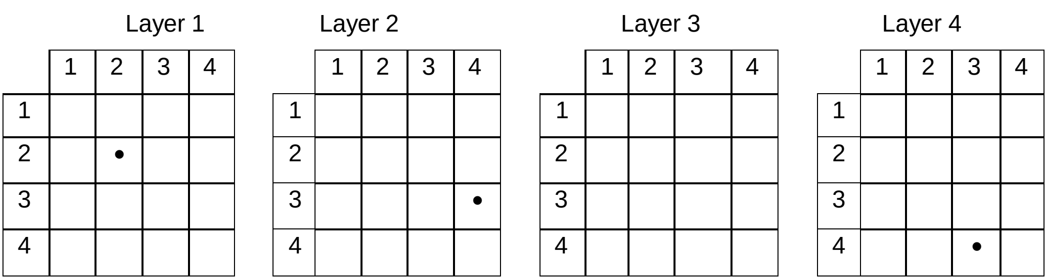

Figure 4. Example of SLCS-1 where d = 4.

5. Estimation of Population Total

Horvitz-Thompson [10] provided an estimator of the population total τ for sampling designs having unequal unit selection probabilities, and sampling is done without replacement from a finite-population which the inclusion probabilities include necessary to determine the estimator variances and the estimators of these variances.

For Simple Latin Cubic Sample + 1 Designs

The first-order inclusion probability π_i is the probability unit u_i will be included by a sampling design. Because every unit has the same probability of being included in a SLCS+1, π_i for i =1, 2,…, d^3 is defined as $\pi_i = (d+1)/d^3$.

The second-order inclusion probability π_{ij} is the probability units u_i and u_j for i, j=1, 2,…, d^3, will both be included by a sampling design. To obtain π_{ij} for SLCS+1 design, three cases need to be considered:

case I: If the sampling units are located in the same row or column or layers, then both units will be included if and only if one of them, say u_i is in the initial SLCS with probability $1/d^2$, and u_j is selected as additional unit with probability $1/(d^3 - d)$ (or vice versa). When u_i and u_j are in the same row or column or layers, the second-order inclusion probability is

$$\pi_{ij} = \frac{2}{d^3(d^2-1)} .$$

case II: If the sampling units u_i and u_j are in different rows, columns and layers, then either (i) both units are in the SLCS, or (ii) one of them is in the SLCS and the other is chosen as addition unit. Since there $[(d-2)!]^2(d^3-d)$ possible samples when u_i and u_j are both members of the SLCS, the probability associated with this case II(i) is $\pi_{ij} = 1/(d^2(d-1)^2)$. There are $2[[(d-1)!]^2 - [(d-2)!]^2]$ possible SLCS in u_i or u_j (but not both) is member of SLCS. Thus, the probability associated with caseII(ii) is $\pi_{ij} = 2(d-2)/(d^2(d-1)^2(d^2-1))$.

Combining the results from case II when u_i and u_j appear in different rows columns and layers, yields the probability,

$$\pi_{ij} = \frac{(d^2+2d-5)}{(d^2(d-1)^2(d^2-1))} .$$

For Simple Latin Cubic Sample -k Designs

If the researcher does not have enough resources and have sample at least d +1 unit. Assume a maximum of d - k (where k = 0, 1,.. , d - 2) units could be sampled. The first-order inclusion probability π_i is $\pi_i = (d-k)/d^3$.

The second order inclusion probability π_{ij} if the sampling units are located in the same row or column or layer, then $\pi_{ij} = 0$. However, when the sampling units are in different rows and columns the second order inclusion probability is

$$\pi_{ij} = \frac{(d-k)(d-k-1)}{d^3(d-1)^3}.$$

When the goal is to estimate τ and the π_i are known, Horvitz and Thompson [10] estimators are used. Simple latin cubic sampling (SLCS+1) has the same Horvitz-Thompson estimator as the two-dimensional simple latin square sampling (SLSS) case.

That is, $\hat{\tau} = \sum_{i=1}^{n} y_i/\pi_i$ when n= 2, 3, …, d, d+1.These estimators are design unbiased estimators of τ where N= d^3 is the size of the population and the summation is over the n units in the sample. The variance of the estimator $\hat{\tau}$ is

$$\text{var}(\hat{\tau}) = \sum_{i=1}^{N}\left(\frac{1}{\pi_i}-1\right)y_i^2 + 2\sum_{i=1}^{N}\sum_{j>1}^{N}\left(\frac{1}{\pi_i\pi_{ij}}-1\right)y_i y_j \qquad .$$

An estimator of this variance is

$$\hat{\text{var}}(\hat{\tau}) = \sum_{i=1}^{n}\left(\frac{1-\pi_i}{\pi_i^2}\right)y_i^2 + 2\sum_{i=1}^{n}\sum_{j>1}^{n}\left(\frac{1}{\pi_i\pi_{ij}}-\frac{1}{\pi_{ij}}\right)y_i y_j .$$

If π_{ij} > 0 for all i, j =1, 2, . . . ,N then the estimator the variance of $\hat{\tau}$ is unbiased in the design sense [3]. Thus, $\hat{\text{var}}(\hat{\tau})$ is an unbiased estimator for a SLCS+1.

6. Simulation Results

In this study, we compared sampling designs when sampling from populations having no or positive spatial correlation. These simulation studies have three populations: no spatial trends (P1), has a strong top-to-bottom (one-dimensional) spatial trend (P2), and has a strong diagonal (two-dimensional) spatial trend (P3).The variances of Horvitz-Thompson estimators of the population total for SLCS+1 and SLCS-*k*, stratified SLCS+1 and SLCS-*k* designs, and systematic SLCS+1 and SLCS-*k* designs were compared, respectively, to the variances of the Horvitz-Thompson estimators for the population for SRS, stratified SRS, and systematic SRS designs. The simulated population data were generated by using MATLAB version 7. Each sampling situation is repeated 100 times and the average variance of the estimator for population total was calculated. The simulation results are reported in Tables 1, 2 and 3.

From Table 1, to evaluate the efficiency of estimators of the population total, the variances of the estimators from SLCS+1 and SLCS-*k* are compared to the variances of the estimators for SRS without replacement. For population sizes $d^3 = 5^3$, 8^3, 10^3, the average of the 100 variances of the estimators were calculated for sample sizes n = 2, 3, .., d, $d+1$ which correspond to the smallest SLCS-*k* (n = 2,3,..d) design to the largest SLCS+1 ($n = d+1$) design.

The reduction in variance depends on type of the spatial correlation that is present in the population. When there is no spatial trend (P1), there are negligible differences between the variances of SLCS+1 or SLCS-*k* estimator and the variance of the SRS estimator. This is an expected result because there is no benefit to having a spatially well-distributed sample in a population exhibiting spatial randomness. If the population has a strong top-to-bottom (one-dimensional) spatial trend (P2) or a strong diagonal (two-dimensional) spatial trend (P3), then the variances of SLCS+1 and SLCS–*k* estimators are smaller than the corresponding SRS estimator variances. Also, the variances of the SLCS estimators are not necessarily a decreasing function of sample size. The variance will decrease from sample size n = 2 to $n = d$, but it may then increase when $n = d + 1$.

The results in Table 2 show that populations were stratified in the following ways: for $d^3 = 8^3$, 8 strata were formed from adjacent cubes composed of $d_h^3 = 4^3$ cubics, for $d^3 = 10^3$, 8 strata were formed from adjacent cubes of $d_h^3 = 5^3$ cubics, and for $d^3 = 20^3$, 6 strata were formed in two different ways: (i) 64 strata were formed from adjacent cubes of $d_h^3 = 5^3$ cubics and (ii) 8 strata were formed from adjacent cubes of $d_h^3 = 10^3$ cubics. For each of these stratified populations, the average variances of the estimators were calculated for sample sizes n_h = 2, 3,…,d_h+1 which correspond to smallest stratified SLCS-*k* to the largest stratified SLCS+1 designs. Also, for each population spatial pattern, the average variances of stratified SLCS-*k* and SLCS+1 design estimators are smaller than the average variances of the corresponding stratified SRS estimator.

Table 1. The average variances of estimators for population total with simple random sampling (SRS) and simple latin cubic sampling (SLCS) designs for three populations.

N	*k*	*n*	The average variances of estimators for population total.					
			P1		P2		P3	
			$\overline{\text{var}}(\hat{\tau}_{srs})$	$\overline{\text{var}}(\hat{\tau}_{slcs})$	$\overline{\text{var}}(\hat{\tau}_{srs})$	$\overline{\text{var}}(\hat{\tau}_{slcs})$	$\overline{\text{var}}(\hat{\tau}_{srs})$	$\overline{\text{var}}(\hat{\tau}_{slcs})$
5^3	-3	2	1.54E+4	1.54E+4	1.71E+4	1.33E+4	9.29E+4	7.40E+4
	-2	3	1.02E+4	1.02E+4	1.13E+4	6.25E+4	6.14E+4	3.62E+4
	-1	4	7.56E+4	7.56E+4	8.42E+4	2.72E+4	4.57E+4	1.73E+4
	0	5	6.00E+4	5.99E+4	6.68E+4	5.96E+4	3.63E+4	5.98E+4
	+1	6	4.96E+4	4.95E+4	5.52E+4	1.37E+4	3.00E+4	9.29E+4
8^3	-6	2	2.63E+4	2.63E+4	7.14E+4	6.17E+4	3.71E+4	3.22E+4
	-5	3	1.75E+4	1.75E+4	4.75E+4	3.45E+4	2.47E+4	1.82E+4
	-4	4	1.31E+4	1.31E+4	3.55E+4	2.10E+4	1.85E+4	1.12E+4
	-3	5	1.05E+4	1.05E+4	2.84E+4	1.29E+4	1.47E+4	6.96E+4
	-2	6	8.70E+4	8.69E+4	2.36E+4	7.43E+4	1.23E+4	4.15E+4
	-1	7	7.44E+4	7.43E+4	2.02E+4	3.55E+4	1.05E+4	2.15E+4
	0	8	6.50E+4	6.49E+4	1.76E+4	6.47E+4	9.15E+4	6.49E+4
	+1	9	5.77E+4	5.76E+4	1.56E+4	2.27E+4	8.12E+4	1.43E+4
10^3	-7	2	1.00E+4	1.00E+4	4.22E+4	3.77E+4	2.16E+4	1.94E+4
	-6	3	6.67E+4	6.67E+4	2.81E+4	2.21E+4	1.44E+4	1.14E+4
	-5	4	5.00E+4	5.00E+4	2.11E+4	1.43E+4	1.08E+4	7.38E+4
	-4	5	3.99E+4	3.99E+4	1.68E+4	9.57E+4	8.63E+4	4.99E+4
	-3	6	3.33E+4	3.32E+4	1.40E+4	6.45E+4	7.18E+4	3.39E+4
	-3	7	2.85E+4	2.85E+4	1.20E+4	4.21E+4	6.15E+4	2.25E+4
	-2	8	2.49E+4	2.49E+4	1.05E+4	2.54E+4	5.38E+4	1.40E+4
	-1	9	2.21E+4	2.21E+4	9.32E+4	1.24E+4	4.77E+4	7.31E+4
	0	10	1.99E+4	1.99E+4	8.38E+4	1.98E+4	4.29E+4	1.99E+4
	+1	11	1.80E+4	1.80E+4	7.61E+4	8.62E+4	3.90E+4	5.22E+4

N = number of population unit in each stratum, *n*= number of sample unit in each stratum

Table 2. Strata formed from adjacent d_h^3=4^3 cubic. The average variances of estimators for population total with stratified random sampling and stratified latin cubic sampling+1 designs for three populations.

d^3 (d_h^3)	k	n_h	The average variances of estimators for population total.					
			P1		P2		P3	
			$\overline{\text{var}}(\hat{\tau}_{st})$	$\overline{\text{var}}(\hat{\tau}_{stslcs})$	$\overline{\text{var}}(\hat{\tau}_{st})$	$\overline{\text{var}}(\hat{\tau}_{stslcs})$	$\overline{\text{var}}(\hat{\tau}_{st})$	$\overline{\text{var}}(\hat{\tau}_{stslcs})$
8^3	-2	2	3.15E+4	3.14E+4	2.34E+4	1.69E+4	1.33E+4	1.00E+4
(4^3)	-1	3	2.06E+4	2.06E+4	1.53E+4	6.63E+4	8.72E+4	4.35E+4
	0	4	1.52E+4	1.51E+4	1.13E+4	1.51E+4	6.43E+4	1.51E+4
	+1	5	1.20E+4	1.19E+4	8.90E+4	2.84E+4	5.06E+4	2.02E+4
10^3	-3	2	1.20E+4	1.20E+4	1.35E+4	1.05E+4	7.35E+4	5.85E+4
(5^3)	-2	3	7.93E+4	7.96E+4	8.93E+4	4.93E+4	4.86E+4	2.86E+4
	-1	4	5.90E+4	5.93E+4	6.64E+4	2.14E+4	3.62E+4	1.37E+4
	0	5	4.68E+4	4.72E+4	5.27E+4	4.72E+4	2.87E+4	4.72E+4
	+1	6	3.87E+4	3.89E+4	4.35E+4	1.08E+4	2.37E+4	7.34E+4
20^3	-3	2	9.80E+4	9.79E+4	1.09E+4	8.46E+4	5.94E+4	4.73E+4
(5^3)	-2	3	6.48E+4	6.46E+4	7.19E+4	3.97E+4	3.93E+4	2.31E+4
	-1	4	4.82E+4	4.80E+4	5.35E+4	1.73E+4	2.92E+4	1.11E+4
	0	5	3.82E+4	3.81E+4	4.24E+4	3.81E+4	2.32E+4	3.81E+4
	+1	6	3.16E+4	3.15E+4	3.51E+4	8.69E+4	1.92E+4	5.93E+4
20^3	-8	2	8.01E+4	8.01E+4	3.38E+4	3.01E+4	1.73E+4	1.55E+4
(10^3)	-7	3	5.34E+4	5.34E+4	2.25E+4	1.76E+4	1.15E+4	9.09E+4
	-6	4	4.00E+4	4.00E+4	1.68E+4	1.14E+4	8.63E+4	5.90E+4
	-5	5	3.20E+4	3.19E+4	1.35E+4	7.65E+4	6.89E+4	3.99E+4
	-4	6	2.66E+4	2.66E+4	1.12E+4	5.15E+4	5.74E+4	2.71E+4
	-3	7	2.28E+4	2.28E+4	9.60E+4	3.37E+4	4.91E+4	1.80E+4
	-2	8	1.99E+4	1.99E+4	8.39E+4	2.03E+4	4.30E+4	1.12E+4
	-1	9	1.77E+4	1.77E+4	7.45E+4	9.91E+4	3.81E+4	5.84E+4
	0	10	1.59E+4	1.59E+4	6.70E+4	1.59E+4	3.43E+4	1.59E+4
	+1	11	1.44E+4	1.44E+4	6.08E+4	6.90E+4	3.11E+4	4.17E+4

n_h = number of sample unit in each stratum, d_h^3= number of population in each stratum

The reduction in average variance depends on the type of spatial correlation. For the populations having no spatial trend (P1), the differences are negligible between the average variances of the estimator for stratified SLCS-*k* or SLCS+1 and the stratified SRS of equal size. For the populations having a strong top-to-bottom (one-dimensional) spatial trend (P2) or a strong diagonal (two-dimensional) spatial trend (P3), the estimators for stratified SLCS-*k* and SLCS+1 designs are more efficient than the corresponding stratified SRS estimators for each sample size. Also, the average variances of the stratified SLCS estimators are not necessarily a decreasing function of within-stratum sample size n_h. The average variance will decrease from $n_h = 2$ to $n_h = d_h$, but it may then increase for $n_h = d_h + 1$.

According to Table 3, consider the common sampling situation when primary units are selected systematic random sampling designs. The simulated populations were systematically partitioned in the following ways: for $d^3 = 8^3$ there were $d_h^3 = 4^3$ primary units each containing 8 secondary units, for $d^3 = 10^3$ there were $d_h^3 = 5^3$ primary units each containing 8 secondary units, for $d^3 = 20^3$ there were two cases: (i) $d_h^3 = 5^3$ primary units each containing 64 secondary units and (ii) $d_h^3 = 10^3$ primary units each containing 8 secondary units.

The reduction in average variance depends on the type of spatial correlation. For the populations having no spatial trend (P1), the differences are negligible between the average variances of the estimator for the systematic SLCS-*k* or SLCS+1 and the systematic SRS of equal size. For the populations having a strong top-to-bottom (one-dimensional) spatial trend (P2) or a strong diagonal (two-dimensional) spatial trend (P3), the estimators for systematic SLCS-*k* and SLCS+1 designs are more efficient than the corresponding systematic SRS estimators for each sample size. Also, the average variances of the systematic SLCS estimators are not necessarily a decreasing function of within-stratum sample size n_h. The average variance will decrease from $n_h = 2$ to $n_h = d_h$, but it may then increase for $n_h = d_h + 1$.

Table 3. The average variances of estimators for population total with systematic random sampling and systematic latin cubic sampling+1 designs for three populations.

d^3 (d_h^3)	k	n_h	The average variances of estimators for population total.					
			P1		P2		P3	
			$\overline{\text{var}}(\hat{\tau}_{sys})$	$\overline{\text{var}}(\hat{\tau}_{sysslcs})$	$\overline{\text{var}}(\hat{\tau}_{sys})$	$\overline{\text{var}}(\hat{\tau}_{sysslcs})$	$\overline{\text{var}}(\hat{\tau}_{sys})$	$\overline{\text{var}}(\hat{\tau}_{sysslcs})$
8^3 (4^3)	-2	2	3.06E+4	3.05E+4	1.65E+4	1.12E+4	8.36E+4	5.76E+4
	-1	3	2.01E+4	2.00E+4	1.08E+4	3.85E+4	5.48E+4	2.02E+4
	0	4	1.48E+4	1.47E+4	7.96E+4	1.47E+4	4.04E+4	1.47E+4
	+1	5	1.16E+4	1.16E+4	6.26E+4	1.43E+4	3.18E+4	7.74E+4
10^3 (5^3)	-3	2	1.23E+4	1.23E+4	1.00E+4	7.62E+4	5.08E+4	3.87E+4
	-2	3	8.15E+4	8.15E+4	6.64E+4	3.41E+4	3.36E+4	1.75E+4
	-1	4	6.06E+4	6.06E+4	4.94E+4	1.31E+4	2.50E+4	6.85E+4
	0	5	4.81E+4	4.80E+4	3.92E+4	4.79E+4	1.98E+4	4.80E+4
	+1	6	3.98E+4	3.97E+4	3.24E+4	5.95E+4	1.64E+4	3.17E+4
20^3 (5^3)	-3	2	1.04E+4	1.04E+4	6.36E+4	4.81E+4	3.19E+4	2.41E+4
	-2	3	6.85E+4	6.84E+4	4.21E+4	2.14E+4	2.11E+4	1.07E+4
	-1	4	5.10E+4	5.08E+4	3.13E+4	8.06E+4	1.57E+4	4.05E+4
	0	5	4.04E+4	4.02E+4	2.48E+4	4.02E+4	1.24E+4	4.02E+4
	+1	6	3.34E+4	3.33E+4	2.05E+4	3.59E+4	1.03E+4	1.81E+4
20^3 (10^3)	-8	2	8.03E+4	8.03E+4	2.65E+4	2.35E+4	1.33E+4	1.18E+4
	-7	3	5.35E+4	5.35E+4	1.76E+4	1.37E+4	8.84E+4	6.90E+4
	-6	4	4.01E+4	4.01E+4	1.32E+4	8.84E+4	6.62E+4	4.44E+4
	-5	5	3.20E+4	3.20E+4	1.05E+4	5.90E+4	5.29E+4	2.97E+4
	-4	6	2.67E+4	2.67E+4	8.78E+4	3.94E+4	4.40E+4	1.98E+4
	-3	7	2.28E+4	2.28E+4	7.52E+4	2.54E+4	3.77E+4	1.28E+4
	-2	8	2.00E+4	1.99E+4	6.57E+4	1.49E+4	3.30E+4	7.53E+4
	-1	9	1.77E+4	1.77E+4	5.84E+4	6.70E+4	2.93E+4	3.44E+4
	0	10	1.59E+4	1.59E+4	5.25E+4	1.59E+4	2.63E+4	1.59E+4
	+1	11	1.45E+4	1.45E+4	4.77E+4	4.51E+4	2.39E+4	2.33E+4

n_h = number of sample unit in each unit, , d_h^3= number of population in each stratum

7. Conclusion and Discussion

In this study we expanded simple latin square sampling designs to three dimension called "simple latin cubic sample +1 designs (SLCS+1)" and "simple latin cubic sample -*k* designs (SLCS-*k*)". Assume that the survey region can be partitioned into a $d \times d \times d$ three-dimensional grid of d^3 equisized three-dimensional rectangular solids which will be referred to as "cubics". The cubics form the population of sampling units. In these sampling designs, the possible sample sizes are n = 2, 3, …, d, d + 1. A SLCS+1 design is composed of two sets of sampling units: (i) d sampling units (cubics) from a SLCS, and (ii) one additional sampling unit selected from the d^3-d remaining cubics. For a SLCS+1 design, one unit is sampled from each the row and column and layer in the population of $d \times d \times d$ of sampling units.

Horvitz-Thompson estimators are unbiased estimators of the population total, and because SLCS+1 and SLCS-*k* have the same first-order inclusion probabilities as SRS, the corresponding Horvitz-Thompson estimators of the population total are equivalent. The estimator of the variance of a Horvitz-Thompson estimator of the population total is unbiased for a SLCS+1 but not for a SLCS-*k* because the second-order inclusion probability for a SLCS-*k* is zero when two sampling units are located in the same row or column or layer.

The results from this study are in agreement with the simple latin square sampling (SLSS) results in Borkowski [3]. That is, for every sample size the variance of the SLCS-*k* and SLCS+1 estimator is smaller than the variance of the SRS estimator. The largest reduction in variance occurred with the population having the strong diagonal trend, followed by population with the north–south trend. For the populations having no spatial trends, the differences between estimator variances was negligible. The same conclusions hold for the estimator variance assuming stratified SLCS-*k* or SLCS+1, or assuming systematic SLCS-*k* and SLCS+1.

Hence, SLCS-*k* and SLCS+1 design, when compared to simple random, stratified, and systematic sampling will provide estimators with smaller variance when spatial correlation exists.

8. Acknowlwdgements

We would like to thank Thammasat University for partial support this research.

References

[1] Cochran, W. G. Sampling Techniques. New York: Wiley, 1977.

[2] Salehi M.M. Adaptive cluster row and column elimination sampling+1 designs. Communications in Statistics: Theory and Methods,2006;35:349–362.

[3] Borkowski, J. J. Simple Latin Square Sampling ± k designs. Communications in Statistics: Theory and Methods, 2003;32:215-237.

[4] Patterson, H.D. The errors of lattice sampling. Journal of the Royal Statistical Society, Ser. B, 1954;16:140–149.

[5] Yates, F. Sampling Methods for Censuses and Surveys. 3rd ed. London: Griffen, 1965;356–362.

[6] Raj, D. Sampling Theory. New York: McGraw-Hill, 1968, 81–84.

[7] Kish, L. Survey Sampling. New York: Wiley, 1965;488–495.

[8] Schneider, M., and Weinrich, B.E. An Abstract Model of Three Dimensional Spatial Data Types. University of Florida. 2004.

[9] Munholland, P.L., and Borkowski, J.J. Simple Latin square sampling+1 designs: A Spatial Designs Using Quadrats. Biometrics,1996;52:125-136.

[10] Horvitz, D.G., and Thompson, D.J. A generalization of sampling without replacement from a finite universe. Journal of the American Statistical Association, 1952;47:663–685.