Thailand Statistician
January 2010; 8(1) : 93-107
www.statassoc.or.th
Contributed paper

k – Tuple Simple Latin Square Sampling Designs

Jurairach Akanisthanon [a], Kamon Budsaba* [a] and John J. Borkowski [b]

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

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

***Author for correspondence**; e-mail: kamon@mathstat.sci.tu.ac.th

Received: 15 September 2009
Accepted: 15 January 2010

Abstract

In this research we present sampling designs that are extensions of simple Latin square sampling (SLSS) designs when the population consists of $N = d \times d$ quadrats. The SLSS designs are extended to larger samples, specifically to samples of size $n = 2d$, $3d$, …, $(d - 1)d$ which we call double SLSS, triple SLSS, etc. The general case for $n = kd$ is called a "*k* – tuple simple Latin square sampling design". The goals are then to derive an estimator of the population total, the true variance of this estimator, and an estimator of this variance. Horvitz – Thompson estimation is used to generate formulae for these three estimation goals. Simulated populations that have different forms of spatial correlation are used to show that the variance and the estimated variance of the estimator of the population for *k* – tuple simple Latin square sampling designs are smaller than the variance and the estimated variance for simple random sampling designs. That is, taking a *k* – tuple SLSS is more efficient than the simple random sampling designs for estimating population total.

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

1. Introduction

In many agricultural, biological, geological and sociological studies, the population of interest is a region that is partitioned into quadrats which represent the sampling units. The variable of interest is some characteristic that can be measured in each quadrat. After taking a sample of these quadrats 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. Moreover, 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 present in many biological and geological populations. Therefore, it is desirable for sampling units to be scattered throughout the population to assure a representative sample with good spatial coverage.

Assume the finite population of interest can be partitioned into a $d \times d$ grid of equal sized rectangular sampling units called quadrats. Simple Latin square sampling (SLSS) designs contain d units such that one sampling unit is selected from each row and each column, and, in general, provides good spatial coverage of the population. In 1996, Munholland and Borkowski [1] considered the following modification: once a simple Latin square sampling of size d is selected, one additional unit is drawn at random from the remaining $d^2 - d$ units to give a new probability sampling design which they called Simple Latin Square Sampling +1 designs (SLSS+1). SLSS+1 designs fall into the classical design framework [2] in that samples generated by the design have corresponding selection probabilities that are independent of the response y. Selection of the one additional unit ensures unbiased variance estimation of a population total τ or population mean μ and also helps to provide estimators that are generally more efficient than those based on simple random sampling (SRS) and for systematic sampling when spatial autocorrelation among units is suspected or known to exist. In conclusion, the SLSS +1 designs generate efficient estimators and can provide good spatial coverage.

In a further extension of SLSS designs to a $d \times d$ grid of quadrats, Borkowski [3] introduced simple Latin square sampling $\pm k$ (SLSS $\pm k$) sampling designs. A SLSS $\pm k$ design contains $d \pm k$ sampling units. For a SLSS - k design, $d - k$ units are selected that are a subset of units from a SLSS design. For a SLSS + k design, d units are selected that form a SLSS design, and then k additional units are selected that span k row and k columns. These designs produce estimators of τ or μ with smaller variance than the estimators based on SRS when the units in the population are spatially correlated.

Without stratification, the only allowable sample sizes for SLSS ± *k* designs are from *2* to *2d-1*.

This study will extend simple Latin square sampling to sample sizes of n = 2*d,* 3*d, …, (d-1)d* taken from this population d^2 quadrats. When $n = 2d$, it is called double simple Latin square sampling (double SLSS); when $n = 3d$, it is called triple SLSS; and so on. The general case for $n = kd$ will be called "*k*–tuple simple Latin square sampling" (*k* – tuple SLSS).

Because of desirable spatial coverage properties possessed by *k* – tuple SLSS designs, they should be considered when spatial correlation is present. These desirable spatial properties lead to more precise estimation of τ or μ than would be achieved by SRS. The new theoretical results that will be developed include determining first-order and second-order inclusion probabilities, estimators of τ and μ, the variance of these estimators, and the estimators of these variances. These inclusion probabilities will be used to derive the Horvitz – Thompson estimators [4]. We will then compare the estimated variances of $\hat{\tau}$ from *k* – tuple SLSS to the estimated variances from SRS. To study the efficiency of *k* – tuple SLSS relative to SRS designs, the variance of the *k* – tuple SLSS Horvitz – Thompson estimator will be compared to the variance of the estimator under SRS.

2. *k* – tuple Simple Latin Square Sampling Designs

This research will extend simple Latin square sampling to larger sample sizes, specifically to sample sizes of $n = 2d, 3d, …., (d – 1)d$ taken from this population of $N = d^2$ quadrats. We will call it Double SLSS when $n = 2d$, Triple SLSS when $n = 3d$, etc. The general case for $n = kd$ will be called " *k – tuple SLSS* ". The new theoretical results that will be developed include determining first – order and second – order inclusion probabilities, estimators of the population total τ or mean μ, and estimators of the variance of the estimator of the population total ($\hat{\text{var}}(\hat{\tau})$). Then we will compare (i) the *k* – tuple SLSS estimator variances ($\text{var}(\hat{\tau})$) with the variances of $\hat{\tau}$ for simple random sampling and (ii) the *k* – tuple SLSS estimator of the variances ($\hat{\text{var}}(\hat{\tau})$) with the variances of $\hat{\tau}$ for simple random sampling.

In this study, we define the methodology for *k* – tuple simple Latin square sampling designs as follows.

2.1 Selecting *k* – tuple Simple Latin Square Sampling Designs

Each sample size is selected from a grid of d^2 quadrats. Let the two dimensions be referred to as "row" and "column", and let (r,c) denote the SLSS unit at row = r and column = c. The following algorithm describes the procedure for selecting the units when the sampling design is a *k* – tuple SLSS.

Step1: Generate a SLSS of size *d*. This will be referred to as the original SLSS. Let $c_1, c_2, \ldots, c_d$ be the columns corresponding to the sampled units in rows 1, 2, …, *d*, respectively, of the original SLSS

Step 2: In row 1, randomly select *k* – 1 units from the remaining *d – 1* units. Let $C_1, C_2, \ldots, C_{k-1}$ be the columns corresponding to the additional *k* – 1 units selected in row 1.

Step 3: Generate *k* – 1 additional SLSSs by cyclically shifting the units of the original SLSS to the right based on the distances between the sampled units in row 1. Specifically, in row r (r = 1, 2,…, *d*) of the ith additional SLSS (*i* = 1, 2, …, *k* – 1) select the unit that is

(i) $A_i = C_i - c_1$ units to the right of c_r if $C_i > c_1$

(ii) $A_i = d + C_i - c_1$ units to the right of c_r if $C_i < c_1$

When necessary, cyclically return to column 1 if shifting takes the sample beyond the *d* th unit in that row.

For the *i*th additional SLSS (*i* = 1,2,…,*k* – 1), Step 3 is mathematically equivalent to selecting unit C_i in row 1, unit $A_i + c_2 \pmod d$ in row 2, … , $A_i + c_d \pmod d$ in row *d,* and if $A_i + c_j \pmod d = 0$ for row j, then the unit in column *d* of row j is selected.

Note that there are $\binom{d}{k}(d-1)!$ possible *k* – tuple SLSS designs.

Consider an example of selecting a *k* – tuple SLSS when *d* = 5 and *k* = 3. We can generate this 3 – tuple SLSS as follows:

Step1: Select the original SLSS. Suppose $(c_1, c_2, c_3, c_4, c_5) = (3, 1, 4, 5, 2)$. Then the SLSS is shown below.

		X		
X				
			X	
				X
	X			

Step 2: Select $k-1=2$ additional units in row 1. Suppose the 2 additional units are in columns $C_1=2$ and $C_2=5$, and are denoted by **1** and **2**, respectively.

	1	X		**2**
X				
			X	
				X
	X			

Step 3: Additional unit 1 is $A_1=4$ units to the right of X in column 3 from the SLSS. Or, mathematically, because $C_1=2$ is less than $c_1=3$ ($C_1<c_1$), the value of $A_1=d+C_1-c_1=5+2-3=4$. Then we move 4 units to the right of the SLSS units in rows 2, 3, 4, and 5. Each unit in the first additional SLSS is shown below as a **1**.

	1	X		2
X				**1**
		1	X	
			1	X
1	X			

Additional unit 2 is $A_2=2$ units to the right of X in column 3 from the SLSS. Or, mathematically, because $C_2=5$ is greater than $c_1=3$ ($C_2>c_1$), the value of $A_2=C_2-c_1=5-3=2$. Then we move 2 units to the right of the SLSS units in rows 2, 3, 4, and 5. Each unit in the second additional SLSS is shown below as a **2.**

	1	X		**2**
X		**2**		1
2		1	X	
	2		1	X
1	X		**2**	

Note that there are $\binom{d}{k}(d-1)!$ possible k – tuple SLSS designs.

2.2 Estimation of Population Total

2.2.1 Inclusion Probabilities for *k* – tuple Simple Latin Square Sampling Designs

In this research study, the first goal is to derive the first – order and second – order inclusion probabilities which will then be used to derive Horvitz – Thompson estimators of τ and μ, the variances of these estimators, and the estimators of these variances.

Because each of the $n=kd$ units in a k – tuple simple Latin square sample has the same probability of being selected, the first-order inclusion probability

$$\pi_i = \frac{kd}{d^2} = \frac{k}{d}.$$

For distinct units u_i and u_j ,the second – order inclusion probability π_{ij} depends on one of the following 3 cases:

(Case I) ***Units u_i and u_j are in the same row*** : There are $\binom{d-2}{k-2}$ possible ways to select the other *k* - 2 units in row 1 that do not correspond to units u_i and u_j, and (d-1)! ways to select remaining units in the original SLSS. Then, for Case I:

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

(Case II) ***Units u_i and u_j are in the same column*** : Units u_i and u_j will be in different SLSS if they are in the same column. Then there are two units in row 1 that correspond to these two SLSS. Like Case 1, there are $\binom{d-2}{k-2}$ possible ways to select the other *k* - 2 units in row 1 that do not correspond to units u_i and u_j, and (d-1)! ways to select remaining units in the original SLSS. Then, for Case II:

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

(Case III) ***Units u_i and u_j are in different rows and columns :*** We have 2 possible cases to consider:

(i) If units u_i and u_j are in different rows and columns *but are in the same SLSS*, there is a corresponding SLSS unit in row 1. Then there are $\binom{d-1}{k-1}$ possible ways to select the other *k* - 1 units in row 1. Because u_i and u_j are in the same SLSS, there are (*d* - 2)! possible ways to select the other *d* – 2 units in that SLSS. Then for Case III(i):

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

(ii) If units u_i and u_j are in different rows and columns *and are in different SLSSs*, we will first select one of the *d* - 2 possible units in the row of u_i to be in the SLSS containing u_i. Then, pick the remaining *k* - 2 units in row 1 that do not correspond to the two SLSSs containing u_i and u_j . Ther are $\binom{d-2}{k-2}$ possible selections. Last, there are (d-2)! possible ways to select the remaining units of the SLSS containing u_i . Then, for Case III(ii):

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

Combining the results from cases III(i) and (ii) when u_i and u_j are in different rows and columns yields $\pi_{ij} = \pi_{ij}^{(i)} + \pi_{ij}^{(ii)}$ or

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

$$= \frac{k^2 d - 2k^2 + k}{d(d-1)^2} = \frac{k(kd-2k+1)}{d(d-1)^2}.$$

2.2.2 The Horvitz – Thompson estimators

Other goals are to derive an estimator of the population total, the true variance of this estimator, and an estimator of this variance. Horvitz – Thompson estimation [4] is used to generate formulas for these three estimation goals.

When the inclusion probabilities for *k* – tuple simple Latin square sampling designs are known, we can use these to estimate the population total (τ), the true variance of this estimator (var($\hat{\tau}$)) and estimator of this variance ($\hat{\text{var}}(\hat{\tau})$).

The Horvitz – Thompson (HT) estimator [4] for *k* – tuple simple Latin square sampling (*k* – tuple SLSS) designs has the same form as the estimator for the two dimensional simple Latin square sampling (SLSS) designs case. The HT estimators

$$\hat{\tau} = \sum_{i=1}^{n} \frac{y_i}{\pi_i} \quad \text{and} \quad \hat{\mu} = \frac{1}{N}\sum_{i=1}^{n} \frac{y_i}{\pi_i}$$

are design unbiased estimators of the population total (τ) and mean (μ) when $N = d^2$ the population size and the summation is over the *n* units in the sample. The variance of the estimator $\hat{\tau}$ is

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

An estimator of this variance is

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

Because $N = d^2$, it follows directly that $\text{var}(\hat{\mu}) = \frac{1}{d^4}\,\text{var}(\hat{\tau})$ and $\hat{\text{var}}(\hat{\mu}) = \frac{1}{d^4}\,\hat{\text{var}}(\hat{\tau})$. If $\pi_{ij} > 0$ for all *i*, *j* = 1, 2, …, *N*, the estimator $\hat{\text{var}}(\hat{\tau})$ is unbiased in the design sense. Thus, $\hat{\text{var}}(\hat{\tau})$ is an unbiased estimator for a *k* – tuple SLSS.

3. *k* – tuple SLSS Examples

To evaluate the efficiency of *k* – tuple SLSS designs, the variance of the *k* – tuple SLSS Horvitz – Thompson estimators will be compared to the variance of the estimators

under simple random sampling (SRS). The SRS estimators can be found in Cochran [2] and Thompson [5]. Four populations exhibiting various levels of spatial correlation will be compared in this study:

22	25	15	12	7	3	4	5
26	31	25	22	17	11	9	17
30	33	26	21	12	11	20	24
26	26	23	18	12	9	19	28
20	21	19	13	12	13	20	22
21	23	15	11	11	10	13	21
19	14	10	10	11	14	9	9
15	12	8	7	4	7	7	9

Figure 1: (Population P1) An 8 x 8 grid with τ = 1019 taken from Munholland and Borkowski [1].

28	27	26	26	28	32	28	33	25	30
23	26	26	1	26	28	26	29	23	27
21	23	24	22	24	25	26	25	24	25
17	20	19	20	22	21	22	21	19	22
15	17	17	17	18	16	18	15	19	18
14	13	14	18	16	13	13	11	14	12
10	9	11	13	11	10	9	8	11	9
8	9	7	10	8	7	6	8	9	7
5	7	6	8	9	6	5	6	5	4
3	4	2	6	5	6	6	4	3	2

Figure 2: (Population P2) A 10 x 10 grid with τ = 1570 and a top to bottom decreasing trend.

18	20	15	20	20	15	19	18	24	23	20	26	29	28	28	31	31	34	28	32
13	20	16	20	15	23	19	26	21	21	24	30	23	26	25	33	31	28	32	38
16	18	20	24	25	26	22	23	26	26	22	27	25	25	34	28	37	36	38	31
17	17	16	22	21	23	22	27	27	24	28	32	29	33	27	37	37	38	35	33
15	19	23	17	21	23	21	23	24	25	31	26	32	34	32	33	31	31	36	37
21	24	20	21	28	26	30	22	31	25	29	29	27	30	29	37	35	32	38	43
23	17	24	25	24	27	31	29	31	34	27	36	29	29	34	39	37	37	40	36
18	24	21	25	27	22	32	32	31	26	28	34	34	37	35	34	38	38	37	40
22	26	28	26	24	29	33	26	27	27	34	31	39	32	36	38	37	40	44	43
23	27	28	29	26	32	25	31	35	34	32	33	37	32	42	40	40	37	42	44
23	21	31	23	30	27	31	30	32	35	30	40	32	37	37	36	40	44	44	40
26	29	31	26	30	31	34	36	30	38	36	32	38	38	37	42	42	41	40	49
28	24	28	27	26	31	32	29	32	33	38	34	39	38	40	37	41	43	42	43
32	25	31	32	29	29	35	38	38	32	36	35	39	42	39	40	44	42	41	45
27	29	35	28	35	35	31	40	35	37	38	44	40	40	47	39	49	48	51	49
30	29	32	32	33	30	36	38	42	36	35	38	44	47	45	49	41	43	44	51
28	35	35	34	34	33	41	33	34	35	39	44	44	48	44	50	49	48	53	54
29	33	32	36	39	33	33	34	35	42	46	47	48	47	46	45	44	52	54	55
28	37	38	37	33	33	34	37	45	40	39	42	42	46	47	48	52	47	46	53
38	39	39	37	34	38	39	45	39	42	45	41	44	51	46	50	52	51	51	53

Figure 3: (Population P3) A 20 x 20 grid with τ = 13354 with an increasing diagonal trend.

1	1	1	1	1	2	1	0	0	0	4	5	0	1	0	1	2	1	0	1
3	2	1	0	1	0	0	0	1	2	2	2	0	2	2	2	0	2	0	1
7	4	1	1	1	1	0	0	0	2	2	0	4	3	2	4	2	1	2	2
0	1	2	0	0	0	0	0	4	6	5	1	5	0	0	0	2	1	2	0
1	1	0	2	3	2	0	0	2	1	3	1	4	1	1	1	2	2	1	1
2	0	0	0	4	3	3	0	1	16	5	0	1	3	8	0	0	1	3	3
0	0	1	14	3	3	1	2	0	8	0	2	0	3	9	0	4	2	1	0
0	0	5	1	8	7	6	6	6	1	0	4	0	0	1	2	2	0	1	2
0	0	2	2	3	2	2	3	1	1	1	3	0	0	2	2	0	3	4	0
0	0	0	0	1	0	3	1	1	1	2	0	2	0	2	0	2	1	1	0
1	8	7	7	8	0	5	0	1	0	1	2	0	0	2	4	2	2	2	4
0	9	1	0	0	1	1	1	0	0	0	1	2	4	0	2	1	3	3	1
0	0	0	1	0	2	4	3	1	2	2	0	0	1	1	2	2	0	2	4
0	1	0	0	1	2	0	2	3	5	2	0	0	2	1	1	2	0	1	3
1	0	0	1	1	0	0	0	2	2	2	1	1	1	0	0	2	0	0	0
0	2	0	2	2	0	1	1	0	2	0	0	1	0	0	1	1	1	5	3
0	0	0	3	2	1	0	0	0	0	0	2	1	0	1	1	1	3	1	2
1	0	0	1	0	3	0	1	0	0	2	1	2	0	0	0	1	1	1	0
0	0	0	0	0	0	0	1	1	1	0	1	0	3	0	2	0	1	1	0
2	0	0	0	0	0	0	0	1	2	0	1	3	0	0	1	0	1	2	4

Figure 4: (Population P4) The 20×20 grid shown in Fig. 4 corresponds to the census data studied by Rathbun and Cressie [6]. This population exhibits a weak spatial correlation [7].

4. Comparison Results

The variances of the new estimator for k – tuple SLSS were compared to the variances of the estimator based on simple random sampling plans using both simulated and real populations. The estimator variance results of this research study for populations P1 to P4 are summarized in Table 1. These tables contain the variance of the estimators using k – tuple SLSS and the SRS for samples of size n = 2d, 3d,…, (d – 1)d. The population size N is 64 for P1, 100 for P2, and 400 for P3 and P4.

Table 1. The true variances of estimators of population total ($\mathrm{var}(\hat{\tau})$) for SRS and k - tuple SLSS designs for populations P1 – P4: $\hat{\tau}_{SRS}$ is the SRS estimator. $\hat{\tau}_{KTUPLE}$ is the k – tuple SLSS estimator.

The true variances of estimators of population total ($\mathrm{var}(\hat{\tau})$)												
k	P1: $d = 8$, τ = 1,019			P2: $d = 10$, τ = 1,570			P3: $d = 20$, τ = 13,354			P4: $d = 20$, τ = 584		
	n	var ($\hat{\tau}_{SRS}$)	var ($\hat{\tau}_{KTUPLE}$)	n	var ($\hat{\tau}_{SRS}$)	var ($\hat{\tau}_{KTUPLE}$)	n	var ($\hat{\tau}_{SRS}$)	var ($\hat{\tau}_{KTUPLE}$)	n	var ($\hat{\tau}_{SRS}$)	var ($\hat{\tau}_{KTUPLE}$)
2	16	10,449.1	2,892.31	20	28,133.33	1,227.65	40	272,162.80	29,852.98	40	13,870.92	12,908.54
3	24	5,805.05	1,606.84	30	16,411.11	716.13	60	171,361.76	18,796.32	60	8,733.54	8,127.60
4	32	3,483.03	964.1	40	10,550.00	460.37	80	120,961.24	13,267.99	80	6,164.85	5,737.13
5	40	2,089.82	578.46	50	7,033.33	306.91	100	90,720.93	9,950.99	100	4,623.64	4,302.85
6	48	1,161.01	321.37	60	4,688.89	204.61	120	70,560.73	7,739.66	120	3,596.16	3,346.66
7	56	497.58	137.73	70	3,014.26	131.53	140	56,160.58	6,160.14	140	2,862.25	2,663.67
8				80	1,758.33	76.73	160	45,360.47	4,975.50	160	2,311.82	2,151.42
9				90	781.48	34.10	180	36,960.38	4,054.11	180	1,883.70	1,753.01
10							200	30,240.31	3,317.00	200	1,541.21	1,434.28
11							220	24,742.07	2,713.91	220	1,260.99	1,173.50
12							240	20,160.21	2,211.33	240	1,027.48	956.19
13							260	16,283.24	1,786.08	260	829.88	772.31
14							280	12,960.13	1,421.57	280	660.52	614.69
15							300	10,080.10	1,105.67	300	513.74	478.09
16							320	7,560.08	829.25	320	385.30	358.57
17							340	5,336.53	585.35	340	271.98	253.11
18							360	3,360.03	368.56	360	171.25	159.36
19							380	1,591.60	174.58	380	81.12	75.49

For population P1 with d = 8 when the sample sizes are n = 16, 24, 32, 40, 48 and 56, the variances are much smaller using k – tuple SLSS designs in comparison to SRS designs. The variances from SRS designs are approximately 3.6 times larger than the variances from k – tuple SLSS designs with the same n.

For population P2 with d = 10 when the sample sizes are n = 20, 30, 40, 50, 60, 70, 80 and 90, the variances are much smaller using k – tuple SLSS designs in comparison to SRS designs. The variances from SRS designs are approximately 22.9 times larger than the variances from k – tuple SLSS designs with the same n.

For population P3 with d = 20 when the sample sizes are n = 40, 60, 80, 100,…, 380, the variances are much smaller using k – tuple SLSS designs in comparison to SRS designs. The variances from SRS designs are approximately 9.1 times larger than the variances from *k* – tuple SLSS designs with the same *n*.

For population P4 of the longleaf pine data with *d* = 20 when the sample sizes are *n* = 40, 60, 80, 100,…, 380, the variances are slightly smaller using *k* – tuple SLSS designs in comparison to SRS designs. The variances from SRS designs are approximately only 1.1 times larger than the variances from *k* – tuple SLSS designs with the same *n*. The reduction in variance is relatively small because the spatial correlation is very weak. For populations P1, P2, and P3, the spatial correlation is much stronger, and that is why *k* – tuple SLSS designs are much more efficient than SRS designs.

For every sample size across all four populations, the variance of the *k* – tuple SLSS estimator is smaller than the variance of the SRS estimator. The reduction in variance depends on both the type and strength of the spatial correlation. That is,

1) For the 10×10 population P2, there is a vertical (north – south) trend and it had the largest reductions in variance.

2) For the 8×8 population P1 and the 20×20 population P3 with strong diagonal trends, there is still a large reduction, but not as large as the reduction in P2.

3) For the 20×20 population P4, there are no trends. Even so, there is a small reduction in the variance using a *k* – tuple SLSS. This is probably due to localized pockets of higher tree counts. That is, a *k* – tuple SLSS is slightly less likely to include multiple high count quadrats than a SRS, thereby reducing the variance.

5. Simulation Results

The estimated variance ($\hat{\text{var}}(\hat{\tau})$) of the estimators for *k* – tuple SLSS are compared to the estimated variances of estimators based on simple random sampling plans using all possible sample sizes for the four populations. The comparisons based on calculated values of the estimated variance ($\hat{\text{var}}(\hat{\tau})$) of the estimators for *k* – tuple SLSS and SRS taken over 1,000 simulated samples for each sample size, population, and sampling method.

For each simulated sample, the estimated variance of the estimator was computed. In this study, we used the Matlab® 7.1 program [8-10] to simulate the sampling procedures and to calculate the estimated variances of the estimators of population total for all sample sizes.

The results of this research study for populations P1 to P4 are summarized in Table 4.2. This table contains the estimated variances of the estimators of population total for a SRS and *k* – tuple SLSS designs for each population P1 to P4 for samples of size $n = 2d, 3d, \ldots, (d-1)d$.

Table 2. Means of the estimated variances of estimators of population total ($\hat{\text{var}}(\hat{\tau})$) for SRS and *k* - tuple SLSS designs for populations P1 – P4: $\hat{\tau}_{SRS}$ is the SRS estimator. $\hat{\tau}_{KTUPLE}$ is the *k* – tuple SLSS estimator.

Means of the estimated variances of estimators of population total ($\hat{\text{var}}(\hat{\tau})$)												
k	**P1: *d* = *8*, τ = 1,019**			**P2: *d* = *10*, τ = 1,570**			**P3: *d* = *20*, τ = 13,354**			**P4: *d* = *20*, τ = 584**		
	n	$\hat{\text{var}}(\hat{\tau}_{SRS})$	$\hat{\text{var}}(\hat{\tau}_{SRS})$	**n**	$\hat{\text{var}}(\hat{\tau}_{SRS})$	$\hat{\text{var}}(\hat{\tau}_{KTUPLE})$	**n**	$\hat{\text{var}}(\hat{\tau}_{SRS})$	$\hat{\text{var}}(\hat{\tau}_{KTUPLE})$	**n**	$\hat{\text{var}}(\hat{\tau}_{SRS})$	$\hat{\text{var}}(\hat{\tau}_{KTUPLE})$
2	16	10,505.83	2,883.07	20	28,014.91	1,287.42	40	270,840.66	28,388.46	40	14,018.15	13,331.33
3	24	5,784.73	1,650.94	30	16,417.08	623.53	60	172,396.22	16,376.30	60	8,669.89	8,240.81
4	32	3,501.96	988.72	40	10,579.02	416.22	80	121,203.01	12,473.87	80	6,124.48	5,731.11
5	40	2,089.28	582.02	50	7,040.99	303.71	100	90,763.26	9,588.66	100	4,669.76	4,338.81
6	48	1,160.23	323.29	60	4,680.49	199.60	120	70,996.81	7,575.28	120	3,609.98	3,367.03
7	56	498.29	137.41	70	3,012.23	130.09	140	55,750.41	5,949.51	140	2,885.25	2,702.26
8				80	1,757.04	77.58	160	45,199.61	4,813.69	160	2,292.96	2,185.14
9				90	782.04	33.80	180	36,826.01	3,920.32	180	1,887.85	1,784.04
10							200	30,279.71	3,191.72	200	1,550.52	1,448.98
11							220	24,808.51	2,619.91	220	1,259.38	1,180.89
12							240	20,215.85	2,143.82	240	1,031.06	960.77
13							260	16,284.01	1,751.08	260	829.53	777.09
14							280	12,938.17	1,402.64	280	663.49	619.16
15							300	10,094.69	1,095.41	300	515.52	482.30
16							320	7,561.31	827.88	320	384.41	360.64
17							340	5,333.49	586.99	340	271.24	254.05
18							360	3,361.29	368.62	360	171.47	159.66
19							380	1,589.59	174.43	380	81.23	75.63

For population P1 with $d = 8$ when the sample sizes are n = 16, 24, 32, 40, 48 and 56, the estimated variances are much smaller using *k* – tuple SLSS designs in comparison to SRS designs when taken 1,000 simulated samples. The estimated variances from SRS designs are approximately 3.6 times larger than the estimated variances from *k* – tuple SLSS designs with the same *n*.

For population P2 with $d = 10$ when the sample sizes are n = 20, 30, 40, 50, 60, 70, 80 and 90, the estimated variances are much smaller using *k* – tuple SLSS designs in comparison to SRS designs when taken 1,000 simulated samples. The estimated variances from SRS designs are approximately 23.1 times larger than the estimated variances from *k* – tuple SLSS designs with the same *n*.

For population P3 with $d = 20$ when the sample sizes are n = 40, 60, 80, 100,…, 380, the estimated variances are much smaller using *k* – tuple SLSS designs in comparison to SRS designs when taken 1,000 simulated samples. The estimated variances from SRS designs are approximately 9.1 times larger than the estimated variances from *k* – tuple SLSS designs with the same *n*.

For population P4 of the longleaf pine data with d = 20 when the sample sizes are n = 40, 60, 80, 100,…, 380, the estimated variances are slightly smaller using k – tuple SLSS designs in comparison to SRS designs when taken 1,000 simulated samples. The estimated variances from SRS designs are approximately only 1.1 times larger than the estimated variances from k – tuple SLSS designs with the same n. The reduction in variance is relatively small because the spatial correlation is very weak. For populations P1, P2, and P3, the spatial correlation is much stronger.

When we want to validate the unbiased property of the estimator of the variance for each sampling design, we can find evidence in Tables 1 and Table 2. We can see that the values of the true variances and mean estimated variances of estimators of the population total almost have equal values. We know that for Horvitz-Thompson estimation [4], E[$\hat{\mathrm{var}}(\hat{\tau})$] = $\mathrm{var}(\hat{\tau})$ for both SRS and k – tuple SLSS designs. Therefore, we expect the average of $\hat{\mathrm{var}}(\hat{\tau})$ should be close to $\mathrm{var}(\hat{\tau})$ when taken over 1,000 simulated samples for each sample size, population, and sampling method.

Although k – tuple SLSS designs are much more efficient than SRS designs (that is, the estimated variance $\hat{\mathrm{var}}(\hat{\tau}_{KTUPLE})$ for a k – tuple SLSS design is smaller than the estimated variance $\hat{\mathrm{var}}(\hat{\tau}_{SRS})$ for a SRS design), it is possible for the estimated variance for a k – tuple SLSS designs to be less than zero ($\hat{\mathrm{var}}(\hat{\tau}_{KTUPLE}) < 0$) especially for small k in populations P1 to P3. This is not a desirable result because we know the true variance must be greater than 0.

6. Conclusions and Discussion

In this study, we extended simple Latin square sampling designs to larger sample sizes, specifically to sample sizes of $n = 2d, 3d, \ldots, (d-1)d$ taken from this population of $N = d^2$ quadrats. These sampling designs with $n = kd$ will be called "*k – tuple simple Latin square sampling (k – tuple SLSS)* " designs. New theoretical results were derived that include determining the first – order and second – order inclusion probabilities used in Horvitz – Thompson estimators of the population total τ or mean μ, the variances of the estimator, and the estimator of the variance of the estimator of the population total. We then compared (i) the k – tuple SLSS estimator variances ($\mathrm{var}(\hat{\tau})$) with the variances of $\hat{\tau}$ for simple random sampling (SRS) and (ii) the k – tuple SLSS estimator of the variances ($\hat{\text{var}}(\hat{\tau})$) with the variances of $\hat{\tau}$ for simple random sampling (SRS) for four populations exhibiting various levels of spatial correlation.

For Horvitz – Thompson estimators [4], we derived the inclusion probabilities that include determining the first – order and the second – order inclusion probabilities. The SRS and *k* – tuple SLSS designs have the same first – order inclusion probabilities but have different second – order inclusion probabilities. Also, these sampling designs have unbiased estimators of the population total and have unbiased estimators of the true variances of the estimators of the population total.

Moreover, we applied *k* – tuple SLSS designs to populations that are spatially correlated that. The variances of estimator of population total for *k* – tuple SLSS designs and simple random sampling designs were calculated and compared. The estimated variances of the estimator of population total for *k* – tuple SLSS designs and for SRS designs were also compared. This comparison was based on calculated statistics from 1,000 simulated samples for each sample size and population.

One practical benefit of this research study was *k* – tuple SLSS expanded the possible sample sizes of SLSS $\pm$ *k* designs from $n = 2,3,\ldots,2d - 1$ to larger sample sizes of size $n = 2d$, $3d$, …., $(d - 1)d$. The researcher now has more choices of design size to implement a sampling design that possesses the desirable spatial properties of simple Latin square sampling designs.

These desirable spatial properties can lead to more precise estimation of a population total or mean than would be achieved by simple random sampling.

7. Acknowledgements

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

References

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

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

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

[4] 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.

[5] Thompson, S.K., *Sampling*. New York, Wiley, 1992.

[6] Rathbun, S.L., and Cressie,N., A space – time survival point process for a longleaf pine forest in southern Georgia, *Journal of the American Statistical Association*, 1994; 89, 1164-1174.

[7] Cressie N., *Statistics for Spatial Data*, New York, Wiley, 1991.

[8] Abdelwahab K, Ronald B.G., *An introduction to numerical method: a MATLAB approach*, 2nd ed, Taylor & Francis Groups, LLC, 2006.

[9] Brian, R., Hunt, R.L., Lipsman, J.M., Rosengerg. *A Guide to MATLAB for Beginners and Experienced Users*, 2nd ed, Cambridge University Press, 2006.

[10] Marc E.H., *Programming in MATLAB*, Brooks/cole, 2002.