Thailand Statistician
July 2008; 6(2) : 241-256
http://statassoc.or.th
Contributed paper

Ratio Estimator Using Two Auxiliary Variables for Adaptive Cluster Sampling

Nipaporn Chutiman* and Bungon Kumphon
Department of Mathematics, Faculty of Science, Mahasarakham University,
Mahasarakham, 44150, Thailand.
***Author for correspondence**; e-mail: nipaporn.c@msu.ac.th

Received: 1 April 2008
Accepted 29 July 2008

Abstract

Chao [1] considered the ratio estimator for the population total of the variable of interest in adaptive cluster sampling by using only one auxiliary variable. In this paper we focus on the use of transformed auxiliary variables for estimating the population total for the variable of interest by using two auxiliary variables (based on Gupta and Shabbir [2]) . Simulations showed the proposed estimator had the smallest estimated mean square error when compared the ratio estimators using only one auxiliary variable.

Keywords: adaptive cluster sampling, auxiliary variable, ratio estimator.

1. INTRODUCTION

Adaptive cluster sampling, proposed by Thompson [4], is an efficient method for sampling rare and hidden clustered populations. In adaptive cluster sampling, an initial sample of units is selected by simple random sampling. If the value of the variable of interest from a sampled unit satisfies a pre-specified condition C, that is $\{i,\ y_i \geq c\}$, then the unit's neighborhood will also be added to the sample. If any other units that are "adaptively" added also satisfy the condition C, then their neighborhoods are also added to the sample. This process is continued until no more units that satisfy the condition are found. The set of all units selected and all neighboring units that satisfy the condition is called a network. The adaptive sample units , which do not satisfy the condition are called edge units. A network and its associated edge units are called a cluster. If a unit is selected in the initial sample and does not satisfy the condition C, then there is only one unit in the network. A neighborhood must be defined such that if unit i is in the neighborhood of unit j then unit j is in the neighborhood of unit i. In this paper, a neighborhood of a unit is defined as the four spatially adjacent units, that is to the left, right, top and bottom of that unit as shown in figure 1.

0	0	0	0	0
0	7	**0**	0	0
0	**0**	2	**0**	0
0	2	4	2	**0**
0	1	5 *	3	**0**
0	**0**	**0**	**0**	9
0	0	0	0	0

Figure 1 The example of network where a unit neighborhood is defined as four spatially adjacent units

Figure 1 illustrates the example of a network. The unit with a star is the initial unit selected. The condition to adaptively added units is a value greater than or equal to 1. Units that are to the left, right, top, and bottom of one another make up a neighborhood. The units in the gray shading form a single network. The units in bold numbers are edge units of the network. The network and its edge units make up a cluster.

In some situations, the researcher obtains observations from more than one variable, including the variable of interest and the auxiliary variable. The use of auxiliary variable is a common method to improve the precision of the estimation of the population parameter. Chao [1] proposed a ratio estimator by using one auxiliary variable in adaptive cluster sampling. In this paper, we will study the ratio estimator by using two auxiliary variables in adaptive cluster sampling. Section 2 and 3 are a review of important concepts of adaptive cluster sampling and the ratio estimation in adaptive cluster sampling, respectively. The proposed estimator using two auxiliary variables is provided in section 4. Simulation studies are in section 5 and the conclusions of this study are in the last section.

2. Adaptive Cluster Sampling

Let y be the variable of interest defined on the finite population, and the population consists of a set of N units $\{u_1, u_2, \ldots, u_N\}$ index by their labels $S = \{1, 2, \ldots, N\}$. With unit i is associated the variable of interest y_i. The population total of y is $\tau_y = \sum_{i=1}^{N} y_i$. Let $\hat{\tau}_{y\_A}$ be the estimator of the population total in adaptive cluster sampling.

Let n denote the initial sample size and ν denote the final sample size. Let ψ_i denote the network that includes unit i and m_i be the number of units in that network. The initial sample of units is selected by simple random sampling without replacement.

The Hansen-Hurwitz estimator of the population total for the variable of interest can be written as (Thompson, [5] ; Thompson and Seber, [6])

$$\hat{\tau}_{y\_A} = \sum_{i=1}^{N} y_i \frac{f_i}{E(f_i)},$$

where f_i is the number of units in the initial sample which fall in the network ψ_i that includes units i. $f_i = 0$ if no units in the initial sample fall in the network ψ_i. Since f_i units are selected from the m_i units in the network ψ_i, f_i has the hypergeometric distribution (N, m_i, n), so $E(f_i) = nm_i/N$, we have

$$\begin{aligned}\hat{\tau}_{y\_A} &= \frac{N}{n} \sum_{i=1}^{N} y_i \frac{f_i}{m_i}, \\ &= \frac{N}{n} \sum_{i=1}^{n} \frac{1}{m_i} \sum_{j \in \psi_i} y_j, \\ &= \frac{N}{n} \sum_{i=1}^{n} (w_y)_i, \qquad (1)\end{aligned}$$

where $(w_y)_i$ is the average of the variable of interest in the network that includes unit i of the initial sample, that is, $(w_y)_i = \frac{1}{m_i} \sum_{j \in \psi_i} y_j$.

The variance of $\hat{\tau}_{y\_A}$ is

$$V\left(\hat{\tau}_{y\_A}\right)=\frac{N\left(N-n\right)}{\left(N-1\right)}\sum_{i=1}^{N}\left(w_i-\frac{\tau_y}{N}\right)^2 . \quad (2)$$

An unbiased estimator of $V\left(\hat{\tau}_{y\_A}\right)$ is

$$\hat{V}\left(\hat{\tau}_{y\_A}\right)=\frac{N\left(N-n\right)}{n\left(n-1\right)}\sum_{i=1}^{n}\left(w_i-\frac{\hat{\tau}_{y\_A}}{N}\right)^2 . \quad (3)$$

3. The Ratio Estimator in Adaptive Cluster Sampling

For adaptive cluster sampling, the initial sample of units is selected by simple random sampling without replacement. Let x be the auxiliary variable. The population total of x is $\tau_x=\sum_{i=1}^{N}x_i$

The modified Hansen-Hurwitz estimator of the population total of the variable of interest can be written as (Thompson, [5] ; Thompson and Seber, [6])

$$\hat{\tau}_{y\_A}=\frac{N}{n}\sum_{i=1}^{n}\left(w_y\right)_i ,$$

The modified Hansen-Hurwitz estimator of the population total of the auxiliary variable (x) can be written as

$$\hat{\tau}_{x\_A}=\frac{N}{n}\sum_{i=1}^{n}\left(w_x\right)_i , \quad (4)$$

where $\left(w_x\right)_i$ is the average of the auxiliary variable in the network that includes unit i of the initial sample, that is, $\left(w_x\right)_i=\frac{1}{m_i}\sum_{j\in\psi_i}x_j$.

Let R be the population ratio between y and x , $R=\frac{\tau_y}{\tau_x}$. The ratio estimator of the population total is

$$\hat{\tau}_{R_x\_A}=\left(\frac{\hat{\tau}_{y\_A}}{\hat{\tau}_{x\_A}}\right)\tau_x$$

$$= \hat{\tau}_{y\_A}\left(\frac{\tau_x}{\hat{\tau}_{x\_A}}\right)$$

$$= \hat{\tau}_{y\_A} + \hat{R}_A\left(\tau_x - \hat{\tau}_{x\_A}\right), \qquad (5)$$

where $\hat{R}_A = \dfrac{\hat{\tau}_{y\_A}}{\hat{\tau}_{x\_A}}$ is the estimator of the population ratio.

The first term in the Taylor's expansion about the point $\left(\tau_x, \tau_y\right)$ gives the approximation

$$\hat{\tau}_{R_x\_A} \approx \hat{\tau}_{y\_A} + R\left(\tau_x - \hat{\tau}_{x\_A}\right)$$

$$\hat{\tau}_{R_x\_A} - \tau_y \approx \hat{\tau}_{y\_A} + R\left(\tau_x - \hat{\tau}_{x\_A}\right) - \tau_y = \hat{\tau}_{y\_A} - R\hat{\tau}_{x\_A}. \qquad (6)$$

So the variance of $\hat{\tau}_{R_x\_A}$ is approximated by

$$V\left(\hat{\tau}_{R_x\_A}\right) \approx E\left(\hat{\tau}_{y\_A} - R\hat{\tau}_{x\_A}\right)^2 = V\left(\hat{\tau}_{y\_A} - R\hat{\tau}_{x\_A}\right). \qquad (7)$$

4. The Proposed Estimator

Let y be the variable of interest and x and z be two auxiliary variables defined on the finite population. Let $\delta_y = \left(\hat{\tau}_{y\_A} - \tau_y\right)/\tau_y$, $\delta_x = \left(\hat{\tau}_{x\_A} - \tau_x\right)/\tau_x$ and $\delta_z = \left(\hat{\tau}_{z\_A} - \tau_z\right)/\tau_z$ then we have $E\left(\delta_y\right) = 0$, $E\left(\delta_x\right) = 0$, $E\left(\delta_z\right) = 0$ and $E\left(\delta_y^2\right) = \lambda A_{y\_A}^2$, $E\left(\delta_x^2\right) = \lambda A_{x\_A}^2$ and $E\left(\delta_z^2\right) = \lambda A_{z\_A}^2$, where

$\hat{\tau}_{j\_A}$ is the estimator of the population total in adaptive cluster sampling, $\left(j = y, x, z\right)$,

$\lambda = N^2\left(N - n\right)/Nn$,

$A_{j\_A}^2 = S_{j\_A}^2/\tau_j^2$,

$$S_{j\_A}^2 = \sum_{i=1}^{N}\left(\left(w_j\right)_i - \frac{\tau_y}{N}\right)^2 \Big/ \left(N-1\right) \quad \left(j = y, x, z\right).$$

Let $\rho_{ij\_A} = S_{ij\_A} / S_{i\_A} S_{j\_A}$ $(i, j = x, y, z$ and $i \neq j)$ be the correlation coefficient between i and j and S_{i_A}, S_{j_A}, S_{ij_A} be the standard deviations and the covariance term.

The modified Hansen-Hurwitz estimator of the population total of the auxiliary variable z can be written as

$$\hat{\tau}_{z\_A} = \frac{N}{n} \sum_{i=1}^{n} (w_z)_i , \tag{8}$$

where $(w_z)_i$ is the average of the auxiliary variable in the network that includes unit i of the initial sample, that is,

$$(w_z)_i = \frac{1}{m_i} \sum_{j \in \psi_i} z_j .$$

Let $u_i = x_i$ and $v_i = z_i$ $(i = 1, 2, \ldots, N)$. The population total of these transformed variable are $U = \sum_{i=1}^{N} x_i = \tau_x$ and $V = \sum_{i=1}^{N} z_i = \tau_z$ .The estimator of U and V are $u = \frac{N}{n} \sum_{i=1}^{n} (w_x)_i = \hat{\tau}_{x\_A}$ and $v = \frac{N}{n} \sum_{i=1}^{n} (w_z)_i = \hat{\tau}_{z\_A}$.

We now introduce the class of ratio estimators. The form of multiplicative generalization for the estimator of a population total can be written as (based on Gupta and Shabbir [2])

$$\hat{\tau}_{R_{xz}} = \hat{\tau}_y \left(f(x) \right)^{J_1} \left(f(z) \right)^{J_2} .$$

So the multiplicative generalization of $\hat{\tau}_{R_{xz}\_A}$ is

$$\hat{\tau}_{R_{xz}\_A} = \hat{\tau}_{y\_A} \left(\frac{\tau_x}{\hat{\tau}_{x\_A}} \right)^{J_1} \left(\frac{\tau_z}{\hat{\tau}_{z\_A}} \right)^{J_2} = \hat{\tau}_{y\_A} \left(\frac{U}{u} \right)^{J_1} \left(\frac{V}{v} \right)^{J_2} . \tag{9}$$

So

$$\hat{\tau}_{R_{xz}\_A} = \tau_y \left(1 + \delta_y\right) \left(1 + \delta_x\right)^{-J_1} \left(1 + \delta_z\right)^{-J_2} , \tag{10}$$

where $\delta_y = \left(\hat{\tau}_{y\_A} - \tau_y\right) / \tau_y$, $\delta_x = \left(\hat{\tau}_{x\_A} - \tau_x\right) / \tau_x$ and $\delta_z = \left(\hat{\tau}_{z\_A} - \tau_z\right) / \tau_z$.

Note 1) If $J_1 = 0$ and $J_2 = 0$ then $\hat{\tau}_{R_{xz}\_A} = \hat{\tau}_{y\_A}$

2) If $(J_1, J_2) = (1,1), (-1,-1), (1,-1)$ and $(-1,1)$ then $\hat{\tau}_{R_{xz}\_A}$ are called ratio-type, product-type, ratio-cum-product-type and product-cum-ratio-type estimator.

A Taylor series expansion is possible in (10) . This gives (based on Gupta and Shabbir [2])

$$\hat{\tau}_{R_{xz}\_A} = \tau_y \Big[1 + \delta_y - J_1\delta_x - J_1\delta_y\delta_x - J_2\delta_z - J_2\delta_y\delta_z + J_1J_2\delta_x\delta_z + \frac{J_1(J_1+1)}{2}\delta_x^2 + \frac{J_2(J_2+1)}{2}\delta_z^2 \Big] + f(\delta_i),$$

where $f(\delta_i)$ represents the third or higher order terms in δ_i's.

Therefore, the estimator of the population total can be approximated by

$$\hat{\tau}_{R_{xz}\_A} \approx \tau_y \left[1 + \delta_y - J_1\delta_x - J_1\delta_y\delta_x - J_2\delta_z - J_2\delta_y\delta_z + J_1J_2\delta_x\delta_z + \frac{J_1(J_1+1)}{2}\delta_x^2 + \frac{J_2(J_2+1)}{2}\delta_z^2 \right]. \quad (11)$$

The expected of the approximation of $\hat{\tau}_{R_{xz}\_A}$ is

$$E\left(\hat{\tau}_{R_{xz}\_A}\right) \approx \tau_y - \tau_y J_1 \lambda \rho_{yx\_A} A_{y\_A} A_{x\_A} - \tau_y J_2 \lambda \rho_{yz\_A} A_{y\_A} A_{z\_A} + \tau_y J_1 J_2 \lambda \rho_{xz\_A} A_{x\_A} A_{z\_A} + \tau_y \frac{J_1(J_1+1)}{2} \lambda A_{x\_A}^2 + \tau_y \frac{J_2(J_2+1)}{2} \lambda A_{z\_A}^2.$$

So the approximate bias is given by

$$B\left(\hat{\tau}_{R_{xz}\_A}\right) \approx \tau_y \lambda \left[\frac{J_1(J_1+1)}{2} A_{x\_A}^2 + \frac{J_2(J_2+1)}{2} A_{z\_A}^2 + J_1 J_2 \rho_{xz\_A} A_{x\_A} A_{z\_A} - A_{y\_A}\left(J_1 \rho_{yx\_A} A_{x\_A} + J_2 \rho_{yz\_A} A_{z\_A} \right) \right] \quad (12)$$

and the approximate MSE is given by

$$MSE\left(\hat{\tau}_{R_{xz}\_A}\right) \approx \tau_y^2 \lambda \Big[A_{y\_A}^2 + J_1^2 A_{x\_A}^2 + J_2^2 A_{z\_A}^2 - 2J_1 \rho_{yx\_A} A_{y\_A} A_{x\_A} - 2J_2 \rho_{yz\_A} A_{y\_A} A_{z\_A} + 2J_1 J_2 \rho_{xz\_A} A_{x\_A} A_{z\_A} \Big]. \tag{13}$$

The value of J_1 and J_2 are derived by minimizing $MSE\left(\hat{\tau}_{R_{xz}\_A}\right)$ in (12) with respect to J_1 and J_2. That is, $\frac{\partial}{\partial J_1} MSE\left(\hat{\tau}_{R_{xz}\_A}\right) = 0$ and $\frac{\partial}{\partial J_2} MSE\left(\hat{\tau}_{R_{xz}\_A}\right) = 0$.

So the optimum values of J_1 and J_2 are

$$J_1 = \frac{A_{y\_A}\left(\rho_{yx\_A} - \rho_{yz\_A}\rho_{xz\_A}\right)}{A_{x\_A}\left(1-\rho_{xz\_A}^2\right)} \quad \text{and} \quad J_2 = \frac{A_{y\_A}\left(\rho_{yz\_A} - \rho_{yx\_A}\rho_{xz\_A}\right)}{A_{z\_A}\left(1-\rho_{xz\_A}^2\right)}. \tag{14}$$

Substituting (14) in (13), we get the minimum MSE of $\hat{\tau}_{R_{xz}\_A}$ as

$$MSE\left(\hat{\tau}_{R_{xz}\_A}\right) \approx \lambda \tau_y^2 A_{y\_A}^2 \left[1 - \frac{\left(\rho_{yx\_A}^2 + \rho_{yz\_A}^2 - 2\rho_{yx\_A}\rho_{yz\_A}\rho_{xz\_A}\right)}{1-\rho_{xz}^2} \right].$$

Because $\frac{\partial^2}{\partial J_1^2} MSE\left(\hat{\tau}_{R_{xz}\_A}\right) \geq 0$ and $\frac{\partial^2}{\partial J_2^2} MSE\left(\hat{\tau}_{R_{xz}\_A}\right) \geq 0$.

Substituting (14) in (12), we get the bias of $\hat{\tau}_{R_{xz}\_A}$ as

$$B\left(\hat{\tau}_{R_{xz}\_A}\right) \approx \frac{\lambda \tau_y A_{y\_A}}{2\left(1-\rho_{xz\_A}^2\right)} \Big[A_{x\_A}\left(\rho_{yx\_A} - \rho_{yz\_A}\rho_{xz\_A}\right) + A_{z\_A}\left(\rho_{yz\_A} - \rho_{yx\_A}\rho_{xz\_A}\right) - A_{y\_A}\left(\rho_{yx\_A}^2 + \rho_{yz\_A}^2 - 2\rho_{yx\_A}\rho_{yz\_A}\rho_{xz\_A}\right) \Big].$$

The J_1 and J_2 were estimated by

$$\hat{J}_1 = \frac{\hat{A}_{y\_A}\left(r_{yx\_A} - r_{yz\_A} r_{xz\_A}\right)}{\hat{A}_{x\_A}\left(1-r_{xz\_A}^2\right)} \quad \text{and} \quad \hat{J}_2 = \frac{\hat{A}_{y\_A}\left(r_{yz\_A} - r_{yx\_A} r_{xz\_A}\right)}{\hat{A}_{z\_A}\left(1-r_{xz\_A}^2\right)},$$

where $\hat{A}^2_{j\_A} = s^2_{j\_A} \Big/ \hat{\tau}^2_j$, $s^2_{j\_A} = \sum_{i=1}^{n} \left(\left(w_j \right)_i - \frac{\hat{\tau}_{y\_A}}{N} \right)^2 \Big/ (n-1)$, $(j = y, x, z)$.

Let $r_{ij\_A} = s_{ij\_A} \big/ s_{i\_A} s_{j\_A}$ $(i, j = x, y, z$ and $i \neq j)$ be the sample correlation coefficient between i and j and s_{i_A} , s_{j_A} , s_{ij_A} be the standard deviation and the covariance term of the sample.

The $MSE\left(\hat{\tau}_{R_{xz}\_A}\right)$ was estimated by

$$M\hat{S}E\left(\hat{\tau}_{R_{xz}\_A}\right) \approx \lambda \hat{\tau}^2_{y\_A} \hat{A}^2_{y\_A} \left[1 - \frac{\left(r^2_{yx\_A} + r^2_{yz\_A} - 2 r_{yx\_A} r_{yz\_A} r_{xz\_A} \right)}{1 - r^2_{xz}} \right].$$

5. Simulation Study

The step of the simulation data X , Y and Z are as follows:

1. To generate the number of network and random the position of the start unit for the network. To generate the value of X , Y and Z from the normal distribution, where $\mu_y = 20$, $\sigma_y = 5$, $\mu_x = 10$, $\sigma_x = 2$ and $\mu_z = 50$, $\sigma_z = 5$.

2. For each the start unit, to consider the unit's neighborhoods and for each unit's neighborhoods , to generate $0 \leq P \leq 1$ if $P \leq 0.3$ then generating X , Y from the bivariate normal distribution with $\rho_{\text{yx}} = 0.80$, where $\mu_y = 5$, $\sigma_y = 2$, $\mu_x = 2$, $\sigma_x = 1$ and generating Y , Z from the bivariate normal distribution with $\rho_{\text{yz}} = 0.80$, where $\mu_z = 10$, $\sigma_x = 3$, otherwise X , Y and Z are equal 0. The values of X , Y and Z are truncated to be the integer. If the values of X , Y and Z are negative values then X , Y and Z are set to be zero.

The populations were shown in Figure 2 – 4.

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	2	24	5	0	0	0	0	0	0	0	0	0	0	0
0	0	1	22	5	4	5	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	2	0	4	8	0	0	0	0	0	33	0	0	0	27	0	0
0	0	0	0	0	0	0	0	0	0	0	0	7	6	7	1	0	5	0	0
0	0	0	0	0	0	0	21	0	0	1	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	4	0	5	7	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	5	7	0	7	7	6	3	0	0	0	0	0	0
0	0	0	5	4	3	0	5	8	4	5	1	0	5	0	0	0	0	0	0
0	7	65	0	4	5	0	9	0	0	0	0	0	3	1	0	0	0	0	0
0	1	4	5	0	7	3	3	0	0	0	0	0	0	0	0	0	0	0	0
0	1	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	27	0	0	21
0	0	0	0	0	0	0	0	0	0	0	0	0	29	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 2. Y values

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	11	3	0	0	0	0	0	0	0	0	0	0	0
0	0	0	11	2	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	2	4	0	0	0	0	0	12	0	0	0	15	0	0
0	0	0	0	0	0	0	0	0	0	0	0	3	2	3	0	0	2	0	0
0	0	0	0	0	0	0	16	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	2	0	2	3	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	2	2	0	3	3	2	1	0	0	0	0	0	0
0	0	0	2	2	1	0	2	3	2	2	0	0	2	0	0	0	0	0	0
0	3	18	0	2	2	0	4	0	0	0	0	0	1	0	0	0	0	0	0
0	0	2	2	0	3	1	1	0	0	0	0	0	0	0	0	0	0	0	0
0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	0	0	12
0	0	0	0	0	0	0	0	0	0	0	0	0	27	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 3. *X* values

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	12	77	8	0	0	0	0	0	0	0	0	0	0	0
0	0	0	57	10	8	14	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	10	0	10	8	0	0	0	0	0	55	1	0	0	97	0	0
0	0	0	0	0	0	0	0	0	0	0	0	17	6	6	1	0	9	0	0
0	0	0	0	0	0	0	74	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	11	0	10	11	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	12	12	0	10	10	12	12	0	0	0	0	0	0
0	0	0	6	8	12	0	12	8	12	10	0	0	10	0	0	0	0	0	0
0	8	53	0	8	6	0	8	0	0	0	0	0	10	0	0	0	0	0	0
0	0	14	10	0	8	12	14	0	0	0	0	0	0	0	0	0	0	0	0
0	0	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	59	0	0	54
0	0	0	0	0	0	0	0	0	0	0	0	0	63	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 4. Z values

The values of the population totals are $\tau_y = 489$, $\tau_x = 222$, $\tau_z = 1012$. The correlation between w_y and w_x is 0.926 and. The correlation between w_y and w_z is 0.949. The values of $S^2_{y\_A}$, $S^2_{x\_A}$, $S^2_{z\_A}$ are 12.686 , 3.794 and 59.834, respectively. The values of $A^2_{y\_A}$, $A^2_{x\_A}$, $A^2_{z\_A}$ are 5.305×10^{-5} , 7.698×10^{-5} and 5.842×10^{-5}, respectively.

For each iteration, an initial sample of units is selected by simple random sampling without replacement. The y-values are obtained for keeping the sample network. In each sample network the x-values and z-values are obtained. We can compute the estimator of the population total of the variable of interest, $\hat{\tau}_{y\_A} = \frac{N}{n}\sum_{i=1}^{n}\left(w_y\right)_i$, the estimator of the population total of the auxiliary variable (x), $\hat{\tau}_{x\_A} = \frac{N}{n}\sum_{i=1}^{n}\left(w_x\right)_i$, and the estimator of the population total of the auxiliary variable (z), $\hat{\tau}_{z\_A} = \frac{N}{n}\sum_{i=1}^{n}\left(w_z\right)_i$. So the multiplicative generalization of $\hat{\tau}_{R_{xz}\_A}$ is

$$\hat{\tau}_{R_{xz}\_A} = \hat{\tau}_{y\_A}\left(\frac{\tau_x}{\hat{\tau}_{x\_A}}\right)^{J_1}\left(\frac{\tau_z}{\hat{\tau}_{z\_A}}\right)^{J_2},$$

For each estimator 5,000 iterations were performed to obtain an accuracy estimate. The condition for added units in the sample is defined by $C = \{y : y > 0\}$. Initial SRS sizes were varied n = *2, 5, 10, 15, 20, 25* and *30* were used.

The estimated mean square error of the estimate total is

$$M\hat{S}E\left(\hat{\tau}\right) = \frac{1}{5,000}\sum_{i=1}^{5,000}\left(\hat{\tau}_i - \tau_y\right)^2 ,$$

where $\hat{\tau}_i$ is the value for the relevant estimator for sample i. Let ν denote the final sample size.

The results of the adaptive cluster sampling $\left(\hat{\tau}_{y\_A}\right)$, using the ratio estimator in adaptive cluster sampling with one auxiliary variable $\left(\hat{\tau}_{R_x\_A}, \hat{\tau}_{R_z\_A}\right)$ and the ratio estimator in adaptive cluster sampling with two auxiliary variables $\left(\hat{\tau}_{R_{xz}\_A}\right)$ are as follows:

Table 1. The estimated bias of the estimators for the population total of the variable of interest.

n	$E(\nu)$	$\hat{B}\left(\hat{\tau}_{R_x\_A}\right)$	$\hat{B}\left(\hat{\tau}_{R_z\_A}\right)$	$\hat{B}\left(\hat{\tau}_{R_{xz}\_A}\right)$
2	13.17	-416.58	-349.49	-355.94
5	29.52	-347.02	-217.78	-251.17
10	54.25	-281.17	-96.59	-158.90
15	70.83	-254.57	-46.15	-121.83
20	87.84	-227.81	-1.08	-85.17
25	99.41	-225.80	1.43	-82.70
30	110.01	-220.50	7.24	-76.81

Table 2. The estimated MSE of the estimators for the population total of the variable of interest.

n	$M\hat{S}E\left(\hat{\tau}_{y\_A}\right)$	$M\hat{S}E\left(\hat{\tau}_{R_x\_A}\right)$	$M\hat{S}E\left(\hat{\tau}_{R_z\_A}\right)$	$M\hat{S}E\left(\hat{\tau}_{R_{xz}\_A}\right)$
2	982,826.38	191,711.49	184,808.59	182,529.52
5	428,116.04	143,203.84	130,289.15	124,493.11
10	225,191.10	94,235.93	70,377.19	64,720.12
15	145,934.65	75,642.93	47,530.53	42,521.95
20	105,545.92	62,212.98	28,899.31	25,656.58
25	88,082.78	56,850.15	21,342.17	19,290.18
30	70,263.95	53,483.88	15,320.62	14,994.71

6. Conclusions

Adaptive cluster sampling is an efficient method for sampling rare and hidden clustered populations. An estimator of the population total in adaptive cluster sampling by using the auxiliary variables can reduce the mean square error of the estimator when the auxiliary variables are correlated with the response of interest. The numerical study showed that the estimated mean square error for the estimators of population total using the auxiliary variable are less than the estimated mean square error for the $\hat{\tau}_{y\_A}$. Moreover the estimated mean square error of the estimator of population total from two auxiliary variables , $\hat{\tau}_{R_{xz}\_A}$, is less than the estimated mean square error of the estimator of population total from only one auxiliary variable , $\hat{\tau}_{R_z\_A}$ and $\hat{\tau}_{R_x\_A}$.

7. Acknowledgements

This research was supported by Research Institute of Mahasarakham University. We would also like to profoundly thanks Mr. Paveen Chutiman for his programming advice.

Appendix A

A.1. Let $\delta_y = \left(\hat{\tau}_{y\_A} - \tau_y\right)/\tau_y$. The expected value is shown as

$$E\left(\delta_y\right) = E\left(\hat{\tau}_{y\_A}/\tau_y\right) - E\left(\tau_y/\tau_y\right)$$

$$= \frac{1}{\tau_y} E\left(\hat{\tau}_{y\_A}\right) - 1$$

$E\left(\hat{\tau}_{y\_A}\right) = \tau_y$ ($\hat{\tau}_{y\_A}$ is unbiased estimator of τ_y)

$$E\left(\delta_y\right) = 0 .$$

Similarly, $E\left(\delta_x\right) = 0$ and $E\left(\delta_z\right) = 0$.

A.2. $E\left(\delta_y^2\right)$ is shown as

$$E\left(\delta_y^2\right) = V\left(\delta_y\right) + \left[E\left(\delta_y\right)\right]^2$$

$$= V\left(\frac{\hat{\tau}_{y\_A} - \tau_y}{\tau_y}\right)$$

$$= \frac{1}{\tau_y^2} V\left(\hat{\tau}_{y\_A}\right)$$

$$= \frac{1}{\tau_y^2} \frac{N^2(N-n)}{Nn} \frac{\sum_{i=1}^{N}\left((w_y)_i - \frac{\tau_y}{N}\right)^2}{(N-1)}$$

$$= \frac{1}{\tau_y^2} \frac{N^2(N-n)}{Nn} S_{y\_A}^2$$

$$E\left(\delta_y^2\right) = \lambda A_{y\_A}^2,$$

where $\lambda = \frac{N^2(N-n)}{Nn}$ and $A_{y\_A}^2 = \frac{S_{y\_A}^2}{\tau_y^2}$.

Similarly, $E\left(\delta_x^2\right) = \lambda A_{x\_A}^2$ and $E\left(\delta_z^2\right) = \lambda A_{z\_A}^2$.

References

[1] Chao, C.T., Ratio Estimation on Adaptive Cluster Sampling: Journal of Chinese Statistical Association. 2004; 42: 307 – 327.

[2] Gupta, S., and Shabbir, J., On the use of Transformed Auxiliary Variables in Estimating Population mean by Using Two Auxiliary Variables: Journal of statistical Planning and Inference. 2007; 137: 1606 – 1611.

[3] Dryver, A.L., Adaptive Sampling Designs and Associated Estimators : Ph.D. Thesis, The Pennsylvania State University,1990.

[4] Thompson, S.K., Adaptive Cluster Sampling: Journal of the American Statistical Association. 1990; 85: 1050 -1059.

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

[6] Thompson, S.K. and Seber, G.A.F., Adaptive Sampling. New York: Wiley,1996.