

## ADAPTIVE CLUSTER SAMPLING USING AUXILIARY VARIABLE

Nipaporn Chutiman

Department of Mathematics, Faculty of Science, Mahasarakham University, Maha Sarakham, Thailand

Received 2013-01-30, Revised 2013-04-02; Accepted 2013-08-21

### ABSTRACT

In this study we study the estimators of the population mean in adaptive cluster sampling by using the information of the auxiliary variable. The estimators in this study are the classical ratio estimator, the ratio estimator using the population coefficient of variation and the coefficient of kurtosis of the auxiliary variable, the regression estimator and the difference estimator. Simulations showed that the difference estimator had the smallest estimated mean square error when compared to the ratio estimators and the regression estimator.

**Keywords:** Adaptive Cluster Sampling, Auxiliary Variable, Ratio Estimator, Regression Estimator, Difference Estimator

### 1. INTRODUCTION

Adaptive cluster sampling, proposed by Thompson (1990), 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 study, 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 **Fig. 1**.

**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.

Sometimes other variables are related to the variable of interest  $y$ . We can obtain additional information for estimating the population mean. Use of an auxiliary variable is a common method to improve the precision of estimates of a population mean. In this study, we will study the estimator of population mean in adaptive cluster sampling using an auxiliary variable. Some comparisons are made using a simulation.

0	0	0	0	0
0	7	<b>0</b>	0	0
0	<b>0</b>	2	<b>0</b>	0
<b>0</b>	2	4	2	<b>0</b>
<b>0</b>	1	5 *	3	<b>0</b>
0	<b>0</b>	<b>0</b>	<b>0</b>	9
0	0	0	0	0

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

### 1.1. Simple Random Sampling Using Auxiliary Variable

Let  $y$  be the variable of interest defined on the finite population and the population consists of a set of  $N$  units  $\{u_1, u_2, \dots, u_N\}$  index by their labels  $S = \{1, 2, \dots, N\}$ . With unit  $i$  is associated the variable of interest  $y_i$  and the auxiliary variable  $x_i$ . The population mean of  $y$  is  $\mu_y = \frac{1}{N} \sum_{i=1}^N y_i$  and the population mean of  $x$  is  $\mu_x = \frac{1}{N} \sum_{i=1}^N x_i$ . Let  $\bar{y}$  be the sample mean of the variable of interest and  $\bar{x}$  be the sample mean of the auxiliary variable in the simple random sampling.

The ratio estimate of the population mean of  $y$  is:

$$\bar{y}_R = \frac{\bar{y}}{\bar{x}} \mu_x = \hat{R} \mu_x$$

Where:

$$\hat{R} = \frac{\bar{y}}{\bar{x}}$$

The approximate MSE of the ratio estimate of the population total of  $y$  is:

$$MSE(\bar{y}_R) \approx \left[ \frac{1-f}{n} \right] (R^2 S_x^2 - 2RS_{xy} + S_y^2)$$

where,  $n$  is the sample size,  $f = \frac{n}{N}$ ,  $R = \frac{\mu_y}{\mu_x}$ , is the population ratio,  $S_x^2$  is the population variance of the auxiliary variable,  $S_y^2$  is the population variance of the variable of interest and  $S_{xy}$  is the population covariance between the auxiliary variable and the variable of interest.

Sisodia and Dwivedi (1981) suggested the ratio estimator for the population mean of  $y$  as:

$$\bar{y}_{R\_SD} = \bar{y} \frac{\mu_x + C_x}{\bar{x} + C_x} = \hat{R}_{SD} (\mu_x + C_x)$$

Where:

$$\hat{R}_{SD} = \frac{\bar{y}}{\bar{x} + C_x}$$

The approximate MSE of  $\bar{y}_{R\_SD}$  is:

$$MSE(\bar{y}_{R\_SD}) \approx \left[ \frac{1-f}{n} \right] \mu_y^2 [C_y^2 + C_x^2 \alpha (\alpha - 2K)]$$

where,  $\alpha = \frac{\mu_x}{\mu_x + C_x}$  and  $K = \rho_{xy} \frac{C_y}{C_x}$ ,  $C_x$  is the population coefficient of variation of the auxiliary variable,  $C_y$  is the population coefficient of variation of the variable of interest and  $\rho_{xy}$  is the coefficient of correlation between the auxiliary variable and the variable of interest.

Singh and Kakran (1993) suggested the ratio estimator for the population mean of  $y$  as:

$$\bar{y}_{R\_SK} = \bar{y} \frac{\mu_x + \beta_2(x)}{\bar{x} + \beta_2(x)} = \hat{R}_{SK} (\mu_x + \beta_2(x))$$

Where:

$$\hat{R}_{SK} = \frac{\bar{y}}{\bar{x} + \beta_2(x)}$$

The approximate MSE of  $\bar{y}_{R\_SK}$  is:

$$MSE(\bar{y}_{R\_SK}) \approx \left[ \frac{1-f}{n} \right] \mu_y^2 [C_y^2 + C_x^2 \delta (\delta - 2K)]$$

where,  $\delta = \frac{\mu_x}{\mu_x + \beta_2(x)}$  and  $\beta_2(x)$  is the population coefficient of kurtosis of the auxiliary variable.

Upadhyaya and Singh (1999) considered both coefficient of variation and kurtosis in ratio estimator as:

$$\bar{y}_{R\_UK1} = \bar{y} \frac{\mu_x \beta_2(x) + C_x}{\bar{x} \beta_2(x) + C_x} = \hat{R}_{UK1} (\mu_x \beta_2(x) + C_x)$$

Where:

$$\hat{R}_{UK1} = \frac{\bar{y}}{\bar{x} \beta_2(x) + C_x}$$

The approximate MSE of  $\bar{y}_{R\_UK1}$  is:

$$MSE(\bar{y}_{R\_UK1}) \approx \left[ \frac{1-f}{n} \right] \mu_y^2 [C_y^2 + C_x^2 \omega_1 (\omega_1 - 2K)]$$

Where:

$$\omega_1 = \frac{\mu_x \beta_2(x)}{\mu_x \beta_2(x) + C_x}$$

$$\bar{y}_{R\_UK2} = \bar{y} \frac{\mu_x C_x + \beta_2(x)}{\bar{x} C_x + \beta_2(x)} = \hat{R}_{UK2} (\mu_x C_x + \beta_2(x))$$

Where:

$$\hat{R}_{UK2} = \frac{\bar{y}}{\bar{x} C_x + \beta_2(x)}$$

The approximate MSE of  $\bar{y}_{R\_UK2}$  is:

$$MSE(\bar{y}_{R\_UK2}) \approx \left[ \frac{1-f}{n} \right] \mu_y^2 [C_y^2 + C_x^2 \omega_2 (\omega_2 - 2K)],$$

Where:

$$\omega_2 = \frac{\mu_x C_x}{\mu_x C_x + \beta_2(x)}$$

The regression estimate of the population mean of y is:

$$\bar{y}_{lr} = \bar{y} + \beta(\mu_x - \bar{x})$$

Where:

$$\beta = S_{xy} / S_x^2$$

The approximate MSE of  $\bar{y}_{lr}$  is:

$$MSE(\bar{y}_{lr}) \approx \left[ \frac{1-f}{n} \right] S_y^2 (1 - \rho_{xy}^2)$$

The difference estimate of the population mean of y is:

$$\bar{y}_D = \bar{y} + (\mu_x - \bar{x})$$

The approximate MSE of  $\bar{y}_D$  is:

$$MSE(\bar{y}_D) \approx \left[ \frac{1-f}{n} \right] (S_y^2 - 2S_{xy} + S_x^2)$$

### 1.2. 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  $\{u_1, u_2, \dots, u_N\}$  index by their labels  $S = \{1, 2, \dots, N\}$ . With

unit i is associated the variable of interest  $y_i$ . The population mean of y is  $\mu_y = \frac{1}{N} \sum_{i=1}^N y_i$ . Let  $\bar{y}_a$  be the estimator of the population total in adaptive cluster sampling.

Let n denote the initial sample size and v denote the final sample size. Let  $\psi_i$  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 mean for the variable of interest can be written as (Thompson, 1990; Thompson and Seber, 1996):

$$\bar{y}_{ac} = \frac{1}{n} \sum_{i=1}^n (w_y)_i$$

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  $\bar{y}_{ac}$  is:

$$V(\bar{y}_{ac}) = \frac{(N-n)}{N(N-1)} \sum_{i=1}^N ((w_y)_i - \mu_y)^2$$

### 1.3. Ratio Estimator in Adaptive Cluster Sampling

Dryver and Chao (2007) proposed the ratio estimator in adaptive cluster sampling as:

$$\bar{y}_{R\_ac} = \frac{\bar{y}_{ac}}{\bar{x}_{ac}} \mu_x$$

Where:

$$\hat{R}_{ac} = \frac{\bar{y}_{ac}}{\bar{x}_{ac}}$$

And:

$$\bar{x}_{ac} = \frac{1}{n} \sum_{i=1}^n (w_x)_i$$

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

$$(w_x)_i = \frac{1}{m_i} \sum_{j \in \psi_i} x_j$$

The approximate MSE of  $\bar{y}_{R\_ac}$  is:

$$MSE(\bar{y}_{R\_ac}) \approx \left[ \frac{1-f}{n} \right] (R^2 S_{wx}^2 - 2RS_{wx-wy} + S_{wy}^2)$$

Where:

$$S_{wx}^2 = \frac{1}{N-1} \sum_{i=1}^N ((w_x)_i - \mu_x)^2$$

$$S_{wy}^2 = \frac{1}{N-1} \sum_{i=1}^N ((w_y)_i - \mu_y)^2$$

And:

$$S_{wx-wy} = \frac{1}{N-1} \sum_{i=1}^N ((w_x)_i - \mu_x)((w_y)_i - \mu_y)$$

### 1.4. Proposed Estimator in Adaptive Cluster

The ratio estimator for the population mean of y in adaptive cluster sampling based on Sisodia and Dwivedi (1981) as:

$$\bar{y}_{R\_ac1} = \bar{y}_{ac} \frac{\mu_x + C_{wx}}{\bar{x}_{ac} + C_{wx}} = \hat{R}_{ac1} (\mu_x + C_{wx})$$

where,  $\hat{R}_{ac1} = \frac{\bar{y}_{ac}}{\bar{x}_{ac} + C_{wx}}$  and  $C_{wx}$  is the population coefficient of  $w_x$ .

MSE of this estimator can be found by using Taylor series method defined as:

$$h(\bar{x}, \bar{y}) \approx h(\mu_x, \mu_y) + \frac{\partial h(\bar{x}, \bar{y})}{\partial \bar{x}} \Big|_{\mu_x, \mu_y} (\bar{x} - \mu_x) + \frac{\partial h(\bar{x}, \bar{y})}{\partial \bar{y}} \Big|_{\mu_x, \mu_y} (\bar{y} - \mu_y)$$

Where:

$$h(\bar{x}, \bar{y}) = \bar{y}_{R\_ac1} = \bar{y}_{ac} \frac{\mu_x + C_{wx}}{\bar{x}_{ac} + C_{wx}}$$

And:

$$h(\mu_x, \mu_y) = \mu_y$$

$$\bar{y}_{R\_ac1} - \mu_y \approx \frac{\partial \left[ \bar{y}_{ac} \frac{\mu_x + C_{wx}}{\bar{x}_{ac} + C_{wx}} \right]}{\partial \bar{x}_{ac}} \Big|_{\mu_x, \mu_y} (\bar{x}_{ac} - \mu_x)$$

$$+ \frac{\partial \left[ \bar{y}_{ac} \frac{\mu_x + C_{wx}}{\bar{x}_{ac} + C_{wx}} \right]}{\partial \bar{y}_{ac}} \Big|_{\mu_x, \mu_y} (\bar{y}_{ac} - \mu_y)$$

$$\approx -\frac{\bar{y}_{ac} (\mu_x + C_{wx})}{(\bar{x}_{ac} + C_{wx})^2} \Big|_{\mu_x, \mu_y} (\bar{x}_{ac} - \mu_x) + \frac{(\mu_x + C_{wx})}{(\bar{x}_{ac} + C_{wx})} \Big|_{\mu_x, \mu_y} (\bar{y}_{ac} - \mu_y)$$

$$\approx \frac{-\mu_y}{(\mu_x + C_{wx})} (\bar{x}_{ac} - \mu_x) + (\bar{y}_{ac} - \mu_y)$$

$$MSE(\bar{y}_{R\_ac1}) = E(\bar{y}_{R\_ac1} - \mu_y)^2$$

$$\approx E \left[ (\bar{y}_{ac} - \mu_y)^2 - \frac{2\mu_y}{(\mu_x + C_{wx})} (\bar{x}_{ac} - \mu_x) (\bar{y}_{ac} - \mu_y) \right.$$

$$\left. + \frac{\mu_y^2}{(\mu_x + C_{wx})^2} (\bar{x}_{ac} - \mu_x)^2 \right]$$

$$\approx V(\bar{y}_{ac}) - \frac{2\mu_y}{(\mu_x + C_{wx})} Cov(\bar{x}_{ac}, \bar{y}_{ac}) + \frac{\mu_y^2}{(\mu_x + C_{wx})^2} V(\bar{x}_{ac})$$

$$\approx \left( \frac{1-f}{n} \right) \mu_y^2 \left[ \frac{S_{wy}^2}{\mu_y^2} + \frac{\mu_x^2 S_{wx}^2}{\mu_x^2 (\mu_x + C_{wx})^2} - \frac{2 \frac{\mu_x}{(\mu_x + C_{wx})} \cdot \frac{\rho_{wx-wy} S_{wx} S_{wy}}{\mu_x \mu_y}}{2} \right]$$

$$\approx \left( \frac{1-f}{n} \right) \mu_y^2$$

$$[C_{wy}^2 + \alpha_w^2 C_{wx}^2 - 2\alpha_w \cdot \rho_{wx-wy} C_{wx} C_{wy}]$$

$$MSE(\bar{y}_{R\_ac1}) \approx \left( \frac{1-f}{n} \right)$$

$$\mu_y^2 [C_{wy}^2 + \alpha_w (\alpha_w - 2K_w)]$$

Where:

$$\alpha_w = \frac{\mu_x}{\mu_x + C_{wx}}$$

And:

$$K_w = \rho_{wx \cdot xy} \frac{C_{wy}}{C_{wx}}$$

The ratio estimator for the population mean of y in adaptive cluster sampling based on Singh and Kakran (1993) as:

$$\bar{y}_{R\_ac2} = \bar{y}_{ac} \frac{\mu_x + \beta_2(w_x)}{\bar{x}_{ac} + \beta_2(w_x)} = \hat{R}_{ac2} (\mu_x + \beta_2(w_x))$$

where,  $\hat{R}_{ac2} = \frac{\bar{y}_{ac}}{\bar{x}_{ac} + \beta_2(w_x)}$  and  $\beta_2(w_x)$  is the population coefficient of kurtosis of  $w_x$ .

The Taylor series method is used for this estimator in the same way to obtain the MSE:

$$MSE(\bar{y}_{R\_ac2}) \approx \left[ \frac{1-f}{n} \right] \mu_y^2 [C_{wy}^2 + C_{wx}^2 \delta_w (\delta_w - 2K_w)]$$

Where:

$$\delta_w = \frac{\mu_x}{\mu_x + \beta_2(w_x)}$$

The ratio estimator for the population mean of y in adaptive cluster sampling based on Upadhyaya and Singh (1999) as:

$$\begin{aligned} \bar{y}_{R\_ac3} &= \bar{y}_{ac} \frac{\mu_x \beta_2(w_x) + C_{wx}}{\bar{x}_{ac} \beta_2(w_x) + C_{wx}} \\ &= \hat{R}_{ac3} (\mu_x \beta_2(w_x) + C_{wx}) \end{aligned}$$

Where:

$$\hat{R}_{ac3} = \frac{\bar{y}_{ac}}{\bar{x}_{ac} \beta_2(w_x) + C_{wx}}$$

and the Taylor series method is used for this estimator in the same way to obtain the MSE:

$$MSE(\bar{y}_{R\_ac3}) \approx \left[ \frac{1-f}{n} \right] \mu_y^2 [C_{wy}^2 + C_{wx}^2 \omega_{w1} (\omega_{w1} - 2K_w)]$$

Where:

$$\omega_{w1} = \frac{\mu_x \beta_2(w_x)}{\mu_x \beta_2(w_x) + C_{wx}}$$

$$\bar{y}_{R\_ac4} = \bar{y}_{ac} \frac{\mu_x C_{wx} + \beta_2(w_x)}{\bar{x}_{ac} C_{wx} + \beta_2(w_x)} = \hat{R}_{ac4} (\mu_x C_{wx} + \beta_2(w_x)),$$

Where:

$$\hat{R}_{ac4} = \frac{\bar{y}_{ac}}{\bar{x}_{ac} C_{wx} + \beta_2(w_x)}$$

and the Taylor series method is used for this estimator in the same way to obtain the MSE:

$$MSE(\bar{y}_{R\_ac4}) \approx \left[ \frac{1-f}{n} \right] \mu_y^2 [C_{wy}^2 + C_{wx}^2 \omega_{w2} (\omega_{w2} - 2K_w)]$$

Where:

$$\omega_{w2} = \frac{\mu_x C_{wx}}{\mu_x C_{wx} + \beta_2(w_x)}$$

The regression estimate of the population mean of y in adaptive cluster sampling is:

$$\bar{y}_{r\_ac} = \bar{y}_{ac} + \beta_w (\mu_x - \bar{x}_{ac})$$

Where:

$$\beta_w = S_{wx \cdot wy} / S_{wx}^2$$

The approximate MSE of  $\bar{y}_{r\_ac}$  is:

$$MSE(\bar{y}_{r\_ac}) \approx \left[ \frac{1-f}{n} \right] S_{wy}^2 (1 - \rho_{wx \cdot wy}^2)$$

The difference estimate of the population mean of y in adaptive cluster sampling is:

$$\bar{y}_{D\_ac} = \bar{y}_{ac} + (\mu_x - \bar{x}_{ac})$$

The approximate MSE of  $\bar{y}_{D\_ac}$  is:

**Table 1.** Data statistics

-----x-variable-----		-----y-variable-----	
$\mu_x$	0.56500	$\mu_y$	1.2275
$S_{wx}^2$	3.79790	$S_{wy}^2$	12.6791
$C_{wx}$	3.44920	$C_{wy}$	2.9008
$\beta_2(w_x)$	92.63470	$\beta_2(w_y)$	23.0357
	5.60070		
$S_{wx \cdot wy}$	0.80710		
$\rho_{wx \cdot wy}$			

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

Fig. 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	2	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	11	2	2	1	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	2	4	0	0	0	0	12	0	0	0	15	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	2	0	2	3	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	2	2	1	0	2	3	2	2	0	0	2	0	0	0	0	0
0	3	18	0	2	2	0	4	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	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	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	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

Fig. 3. X values

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

n	$M\hat{S}E(\bar{y}_{ac})$	$M\hat{S}E(\bar{y}_{R\_ac})$	$M\hat{S}E(\bar{y}_{R\_ac1})$	$M\hat{S}E(\bar{y}_{R\_ac2})$	$M\hat{S}E(\bar{y}_{R\_ac3})$	$M\hat{S}E(\bar{y}_{R\_ac4})$	$M\hat{S}E(\bar{y}_{lr\_ac})$	$M\hat{S}E(\bar{y}_{D\_ac})$
5	1.3511	0.9678	1.0944	1.3323	1.3490	1.2917	0.9717	0.5769
10	0.8560	0.6551	0.7694	0.8502	0.6396	0.8373	0.7123	0.4150
15	0.6698	0.4998	0.6352	0.6678	0.4916	0.6633	0.5638	0.3639
20	0.5855	0.4084	0.5615	0.5843	0.4186	0.5814	0.4685	0.3374
30	0.4999	0.3357	0.4823	0.4991	0.3533	0.4971	0.3664	0.2933
40	0.4488	0.2986	0.4337	0.4481	0.3133	0.4466	0.3071	0.2558

$$MSE(\bar{y}_{D-ac}) \approx \left[ \frac{1-f}{n} \right] (S_{wy}^2 - 2S_{wx-wy} + S_{wx}^2)$$

### 1.5. Simulation Study

In this section, the simulation x-values and y-values from Pochai (2008) were studied. The populations were shown in **Fig. 2-3** and the data statistics of this populations were shown in **Table 1**.

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 the sample network the x-values are obtained. The condition for added units in the sample is defined by  $C = \{y: y>0\}$ .

For each estimator 5,000 iterations were performed to obtain an accuracy estimate. Initial SRS sizes were varied  $n = 5, 10, 15, 20, 30$  and  $40$  were used. The estimated mean square error of the estimate mean is:

$$M\hat{S}E(\bar{y}) = \frac{1}{5,000} \sum_{i=1}^{5,000} (\bar{y}_i - \mu_y)^2$$

where,  $\bar{y}_i$  is the value for the relevant estimator for sample  $i$ .

## 2. CONCLUSION

Adaptive cluster sampling is an efficient method for sampling rare and hidden clustered populations. From the estimated MSE of the estimators in **Table 2** showed that the difference estimator had the smallest estimated mean square error when compared to the ratio estimators and the regression estimator.

The ratio estimator for the population mean ( $\bar{y}_{R-ac}$ ) had the smaller estimated mean square error when compared to the ratio estimators using information of  $C_{wx}$  and  $\beta_2(w_x)$ .

The estimator for the population mean did not use auxiliary variable had the higher estimated mean square error when compared to the estimator for the population mean using auxiliary variable

## 3. ACKNOWLEDGEMENT

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

## 4. REFERENCES

- Dryver, A.L. and C. Chao, 2007. Ratio estimators in adaptive cluster sampling. *Environmetrics*, 18: 607-620. DOI: 10.1002/env.838
- Pochai, N., 2008. Ratio estimator using two auxiliary variables for adaptive cluster sampling. *J. Thai Statist. Assoc.*, 6: 241-256.
- Singh, H.P. and M.S. Kakran, 1993. A modified ratio estimator using known coefficient of kurtosis of an auxiliary character.
- Sisodia, B.V.S. and V.K. Dwivedi, 1981. A modified ratio estimator using coefficient of variation of auxiliary variable. *J. Indian Soc. Agric. Statist.*, 33: 13-18.
- Thompson, S.K. and G.A.F. Seber, 1996. *Adaptive Sampling*. 1st Edn., Wiley, New York, ISBN-10: 0471558710, pp: 265.
- Thompson, S.K., 1990. Adaptive cluster sampling. *J. Am. Statist. Assoc.*, 85: 1050-1059. DOI: 10.1080/01621459.1990.10474975
- Upadhyaya, L.N. and H.P. Singh, 1999. Use of transformed auxiliary variable in estimating the finite population mean. *Biometri. J.*, 41: 627-636. DOI: 10.1002/(SICI)1521-4036(199909)41:5<627::AID-BIMJ627>3.0.CO;2-W