Songklanakarin J. Sci. Technol.
47 (3), 192–199, May – Jun. 2025

http://www.sjst.psu.ac.th

Original Article

Ratio estimators under inverse sampling

Prayad Sangngam and Wipawan Laoarun*

Department of Statistics, Faculty of Science,
Silpakorn University, Mueang, Nakhon Pathom, 73000 Thailand

Received: 4 July 2024; Revised: 18 April 2025; Accepted: 18 April 2025

Abstract

In using a fixed sample size sampling design for a rare population, a challenge encountered lies in the possible production of an estimated population mean of zero by the sample. Inverse sampling is an efficient sampling method for parameter estimation in a rare population. In this study, the covariance of Murthy estimates and its unbiased estimator are derived under inverse sampling. Two ratio estimators of the population mean are proposed. The properties of these ratio estimators are compared to an unbiased estimator of the population mean using a simulation study. Results indicate that the ratio estimators exhibit higher efficiency compared to the unbiased estimator. Moreover, the biased size of the ratio estimator decreases when the number of units of interest in the sample increases. Furthermore, the mean square error of the estimators decreases when the number of units of interest in the sample increases. The efficiency of the two ratio estimators increases when the correlation coefficient between the study and auxiliary variable increases. The proposed estimators are applied to estimate yield of off-season rice in Thailand.

Keywords: unbiased estimator, ratio estimators, inverse sampling, mean square error

1. Introduction

A rare population refers to a demographic group comprising only a small number of units that exhibit the characteristics of interest. A problem encountered in fixed sample size sampling for the population lies in the production of a zero estimate by the sample for the population mean or total. In surveys targeting rare populations, the units can be classified into two groups: those belonging to the class of interest and the remaining units. The class to which a unit belongs is assumed to remain unknown until the unit is observed. Thus, inverse sampling can be regarded as an efficient sampling methodology for estimating the parameters of such populations. In inverse sampling, units are drawn until a fixed number of units exhibiting the characteristics of interest is realized.

Haldane (1945) introduced an inverse sampling approach employing equal probability with replacement. An unbiased estimator for the proportion of units of interest, along with its variance, was derived. However, no unbiased estimator was provided for the variance itself. Finney (1949) derived an unbiased estimator for the variance. Chistman and Lan (2001) explored inverse sampling with and without replacement, involving the selection of units with equal probability in each draw. They introduced an unbiased estimator for the population total and its variance, but failed to provide a specific unbiased estimator for the variance. Salehi and Seber (2001) proved that Murthy's estimator can be adapted for inverse sampling designs. Employing this method, they derived an unbiased variance estimator for the estimator proposed by Christman and Lan (2001). An inverse sampling method with unequal probabilities with replacement was developed to enhance estimation efficacy during the design stage, yielding higher efficiency compared to sampling with equal probabilities in certain conditions (Greco & Naddeo, 2007). Sangngam and Suwattee (2012) derived unbiased estimators for the population mean and its variance under a stratified inverse sampling design.

When one or more auxiliary variables are associated with the study variable, using useful auxiliary information is a way to improve the quality of the estimators. A ratio estimator, which incorporates auxiliary information, is utilized during the estimation stage. In simple random sampling,

*Corresponding author
Email address: laoarun_w@su.ac.th

Murthy (1964) proposed that the ratio estimator outperforms the unbiased estimator when the correlation coefficient between the auxiliary and study variables is large. Sangngam (2014) presented ratio estimators of population mean under simple and stratified random sampling using the coefficients of variation and correlation from the study and auxiliary variables. Sungsuwan and Suwattee (2014) developed an estimator for the total population using the model-assisted approach within the inverse sampling framework. The results indicated that in instances where the correlation coefficient between the auxiliary and study variables is high, the model-assisted estimator exhibits notably greater efficiency compared to the unbiased estimator. Sharma and Kumar (2021) introduced a class of ratio-cum-product type estimators for estimating the population mean using an auxiliary variable in two-phase sampling.

Estimators will be examined in this study using the conventional approach. This study aims to introduce ratio estimators for estimating the population mean under inverse random sampling. The properties of these ratio estimators are derived, and a simulation study is conducted to compare their properties with those of the unbiased estimator. The proposed ratio estimates are applied for estimating the yield of off-season rice of Thailand in 2021.

2. Materials and Methods

2.1 Conventional estimators

Suppose that a finite population comprises N distinct units labeled 1, 2,…, N and is associated with study and auxiliary values (y_1, x_1), (y_2, x_2),…, (y_n, x_n). The parameter of interest is the population mean:

$$\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i$$

Under simple random sampling without replacement, an unbiased estimator of the population mean is represented by the sample mean:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \,.$$

The variance of the unbiased estimator is

$$V(\bar{y}) = \frac{1-f}{n} S_y^2,$$

where $f = \frac{n}{N}$ and $S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(y_i - \bar{Y}\right)^2$.

A typical ratio estimator is

$$\bar{y}_R = \bar{y}\frac{\bar{X}}{\bar{x}},$$

where $\bar{X}$ and $\bar{x}$ are the population and sample means of the auxiliary variable, respectively.

The efficiency of the ratio estimator depends on the coefficient of variation of the auxiliary variable (C_X) and the coefficient of variation of the study variable (C_Y). According to Murthy (1964), if $\rho > \frac{C_X}{2C_Y}$ and the correlation coefficient is positive, then the ratio estimator outperforms the unbiased estimator where ρ is the correlation coefficient between x and y.

The approximate bias and mean square error (MSE) of the ratio estimator are as follows (Cochran, 1977, p. 154):

$$B(\bar{y}_R) = \frac{1-f}{n}\left(\frac{R}{\bar{X}} S_X^2 - \frac{1}{\bar{X}} S_{XY}\right),$$

$$MSE(\bar{y}_R) = \frac{1-f}{n}\left(S_Y^2 + R^2 S_X^2 - 2RS_{XY}\right),$$

where $R = \frac{\bar{Y}}{\bar{X}}$, $S_X^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(x_i - \bar{X}\right)^2$, and

$$S_{XY} = \frac{1}{N-1}\sum_{i=1}^{N}\left(y_i - \bar{Y}\right)\left(x_i - \bar{X}\right) .$$

The sample estimate of the MSE, $M\hat{S}E(\bar{y}_R)$, is calculated as follows:

$$M\hat{S}E(\bar{y}_R) = \frac{1-f}{n}\left(\hat{S}_Y^2 + \hat{R}^2\hat{S}_X^2 - 2\hat{R}\hat{S}_{XY}\right),$$

where $\hat{S}_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$, $\hat{S}_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2$,

$\hat{S}_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)\left(x_i - \bar{x}\right)$, and $\hat{R} = \frac{\bar{y}}{\bar{x}}$.

Assume the division of population units into two classes according to whether the study values satisfy a condition. A common form of the condition is {y : $y > c$}, where c is a given constant. Class C is the class of units in which the study values satisfy the condition, while C' is the class of the remaining units. The cardinalities of classes C and C' are M and N - M, respectively.

In inverse random sampling, units are individually drawn with equal probabilities and without replacement until the sample comprises m units from the class C. Let the sample size be denoted by n. The sample s can be partitioned into two parts: part s_C is the set of sample units from class C and $s_{C'}$ is the set of sample units from C' with cardinalities m and n-m, respectively, where $s_C \cap s_{C'} = \phi$ and $s_C \cup s_{C'} = s$. Under inverse random sampling, an unbiased estimator of $\bar{Y}$ is provided by the following:

$$\bar{y}_M = \hat{P}\bar{y}_C + (1-\hat{P})\bar{y}_{C'},$$

where $\hat{P} = \frac{m-1}{n-1}$, $\bar{y}_C = \frac{1}{m}\sum_{i\in s_C} y_i$, and $\bar{y}_{C'} = \frac{1}{n-m}\sum_{i\in s_{C'}} y_i$

This estimator is Murthy's estimator, which was derived by Salehi and Seber (2001). The "M" subscript denotes that the quantity is a Murthy's estimator. The variance of $\bar{y}_M$ is given by

$$V(\bar{y}_M) = \left(\bar{Y}_C - \bar{Y}_{C'}\right)^2 V\left(\hat{P}\right) + \left(1 - \frac{m}{M}\right)\frac{S_{YC}^2}{m} E\left(\hat{P}^2\right) + \frac{S_{YC'}^2}{m-1} E\left[\hat{P}\left(1-\hat{P}\right)\left(1 - \frac{n-m}{N-M}\right)\right],$$

where $V(\hat{P})$ is the variance of estimator $\hat{P}$ for estimating the parameter $P = \frac{M}{N}$, $S_{YC}^2 = \frac{1}{M-1}\sum_{i\in C}\left(y_i - \bar{Y}_C\right)^2$, $S_{YC'}^2 = \frac{1}{N-M-1}\sum_{i\in C'}\left(y_i - \bar{Y}_{C'}\right)^2$, $\bar{Y}_C = \frac{1}{M}\sum_{i\in C} y_i$, and $\bar{Y}_{C'} = \frac{1}{N-M}\sum_{i\in C'} y_i$.

An unbiased estimator of the variance is

$$\hat{V}(\bar{y}_M) = (\bar{y}_C - \bar{y}_{C'})^2 \hat{V}(\hat{P}) + \frac{\hat{S}_{YC}^2}{m}\left(\hat{P}^* - \frac{m}{N}\hat{P}\right) + \hat{S}_{YC'}^2\left(\frac{\hat{P}(1-\hat{P})}{m-1} - \frac{1-\hat{P}}{N} - \frac{\hat{V}(\hat{P})}{n-m}\right),$$

where $\hat{V}(\hat{P}) = \frac{\hat{P}(1-\hat{P})}{n-2}$, $\hat{P}^* = \frac{(m-1)(m-2)}{(n-1)(n-2)}$, $\hat{S}_{YC}^2 = \frac{1}{m-1}\sum_{i\in s_C}(y_i - \bar{y}_C)^2$, and $\hat{S}_{YC'}^2 = \frac{1}{n-m-1}\sum_{i\in s_{C'}}(y_i - \bar{y}_{C'})^2$.

All of the formulas in this study can be developed similarly for the auxiliary variable, *x*, to derive the proposed ratio estimators.

2.2 Simulation study

Smith *et al.* (1995) provided the data on ring-necked ducks, which was used as the study population. This population is rare and clustered. The population comprises N = 200 units. The number of ring-necked ducks is used as the study variable (y). The true value of the population mean is equal to $\bar{Y} = 116.665$. Auxiliary variables (x's) correlated to the study variable are created with correlation coefficients (ρ) equal to 0.1, 0.3, 0.5, 0.7, and 0.9. Simulations of sampling from this population were conducted to examine the properties of the proposed ratio estimators compared to the unbiased estimator. The condition $\{y : y > 0\}$ was selected for dividing the units into class C or C'. Inverse random sampling was employed to draw the population units. The numbers of units of interest in the sample (m) are 3, 4, 5, 6, and 7. The estimates of bias and MSE of the proposed ratio estimators were compared to those of the unbiased estimator.

3. Results and Discussion

3.1 Proposed ratio estimators

Under inverse random sampling, two ways are available for constructing a ratio estimator of the population mean. The first ratio estimator is derived from the unbiased estimators of the population mean of the study and auxiliary variables. The ratio estimator of $\bar{Y}$ is given by

.

$$\bar{y}_{R1} = \frac{\bar{y}_M}{\bar{x}_M}\bar{X} \quad . \tag{1}$$

The estimator requires a knowledge of the entire population mean of the auxiliary variable, $\bar{X}$.

In order to find the approximate mean square error of $\bar{y}_{R1}$, it can be represented that the covariance of $\bar{y}_M$ and $\bar{x}_M$ is

$$Cov(\bar{y}_M, \bar{x}_M) = \left(\bar{X}_C - \bar{X}_{C'}\right)\left(\bar{Y}_C - \bar{Y}_{C'}\right)V(\hat{P}) + \left(1 - \frac{m}{M}\right)\frac{S_{XYC}}{m}E\left(\hat{P}^2\right) + \frac{S_{XYC'}}{m-1}E\left[\hat{P}\left(1-\hat{P}\right)\left(1 - \frac{n-m}{N-M}\right)\right] \tag{2}$$

where $S_{XYC}^2 = \frac{1}{M-1}\sum_{i\in C}\left(x_i - \bar{X}_C\right)\left(y_i - \bar{Y}_C\right)$ and $S_{XYC'}^2 = \frac{1}{N-M-1}\sum_{i\in C'}\left(x_i - \bar{X}_{C'}\right)\left(y_i - \bar{Y}_{C'}\right)$.

Proof of (2): Apply the variance of $\bar{y}_M$ to the variate $u_i = y_i + x_i$. The population mean of u is $\bar{U} = \bar{Y} + \bar{X}$, and the variance of the estimator, $\bar{u}_M = \bar{x}_M + \bar{y}_M$ of $\bar{U}$, is

$$E\left(\bar{u}_M - \bar{U}\right)^2 = \left(\bar{U}_C - \bar{U}_{C'}\right)^2 V(\hat{P}) + \left(1 - \frac{m}{M}\right)\frac{S_{UC}^2}{m}E\left(\hat{P}^2\right) + \frac{S_{UC'}^2}{m-1}E\left[\hat{P}\left(1-\hat{P}\right)\left(1 - \frac{n-m}{N-M}\right)\right] \tag{3}$$

The quadratic terms are expanded as follows:

$$\begin{aligned} E\left(\bar{u}_M - \bar{U}\right)^2 &= E\left[\left(\bar{y}_M - \bar{Y}\right) - \left(\bar{x}_M - \bar{X}\right)\right]^2 \\ &= E\left(\bar{y}_M - \bar{Y}\right)^2 + E\left(\bar{x}_M - \bar{X}\right)^2 \\ &\quad + 2E\left(\bar{y}_M - \bar{Y}\right)\left(\bar{x}_M - \bar{X}\right), \\ &= V\left(\bar{y}_M\right) + V\left(\bar{x}_M\right) + 2Cov\left(\bar{y}_M, \bar{x}_M\right) \end{aligned}$$

$$\begin{aligned} \left(\bar{U}_C - \bar{U}_{C'}\right)^2 &= \left[\left(\bar{X}_C - \bar{X}_{C'}\right) + \left(\bar{Y}_C - \bar{Y}_{C'}\right)\right]^2 \\ &= \left(\bar{X}_C - \bar{X}_{C'}\right)^2 + \left(\bar{Y}_C - \bar{Y}_{C'}\right)^2 + 2\left(\bar{X}_C - \bar{X}_{C'}\right)\left(\bar{Y}_C - \bar{Y}_{C'}\right), \end{aligned}$$

$$S_{UC}^2 = S_{YC}^2 + S_{XC}^2 + 2S_{XYC},$$

and $$S_{UC'}^2 = S_{YC'}^2 + S_{XC'}^2 + 2S_{XYC'}.$$

Substituting $E\left(\bar{u} - \bar{U}\right)^2$, $\left(\bar{U}_C - \bar{U}_{C'}\right)^2$, S_{UC}^2 and $S_{UC'}^2$ in equation (3), the variance $V\left(\bar{y}_M\right)$ with a similar relation $V\left(\bar{x}_M\right)$ will be canceled. Therefore, the result of Equation (2) follows form the cross-product terms.

To find an estimator of $MSE\left(\bar{y}_{R1}\right)$, the unbiased estimator of (2) will be derived. An unbiased estimator of the covariance is given by

$$C\hat{o}v(\bar{y}_M, \bar{x}_M) = (\bar{x}_C - \bar{x}_{C'})(\bar{y}_C - \bar{y}_{C'})\hat{V}(\hat{P}) + \frac{\hat{S}_{XYC}}{m}\left(\hat{P}^* - \frac{m}{N}\hat{P}\right) + \hat{S}_{XYC'}\left(\frac{\hat{P}(1-\hat{P})}{m-1} - \frac{1-\hat{P}}{N} - \frac{\hat{V}(\hat{P})}{n-m}\right), \tag{4}$$

where $\hat{S}_{XYC} = \frac{1}{m-1}\sum_{i \in s_C}(x_i - \bar{x}_C)(y_i - \bar{y}_C)$ and

$$\hat{S}_{XYC'} = \frac{1}{n-m-1}\sum_{i \in s_{C'}}(x_i - \bar{x}_{C'})(y_i - \bar{y}_{C'}).$$

Proof of (3): Given the sample size, *n*, let E_1 and E_2 be the unconditional and conditional expectations, respectively. Considering the sample size, the inverse random sampling scheme is similar to the stratified random sampling with two strata, *C* and *C'*, where the *m* and *n* - *m* units are selected form the first and second strata, respectively. In addition, the samples from two strata are independent (Greco and Naddeo, 2007).

$$E_1(\hat{P}) = P,\ E_1(\hat{P}^*) = P^2,\ E_2(\bar{y}_C \mid n) = \bar{Y}_C,$$
$$E_2(\bar{y}_{C'} \mid n) = \bar{Y}_{C'},\ E_2(\hat{S}_{XYC} \mid n) = S_{XYC},$$
$$E_2(\hat{S}_{XYC'} \mid n) = S_{XYC'}, \text{ and } E_2(\bar{y}_C \bar{y}_{C'} \mid n) = \bar{Y}_C \bar{Y}_{C'}.$$

Taking the expectation of (3) and using algebra, the unbiased property is proven.

Using a Taylor series approximately, so that the terms of degrees greater than two are ignored, the bias of $\bar{y}_{R1}$ is given by

$$B(\bar{y}_{R1}) = R\frac{V(\bar{x}_M)}{\bar{X}} - \frac{Cov(\bar{x}_M, \bar{y}_M)}{\bar{X}}, \quad (5)$$

where $R = \frac{\bar{Y}}{\bar{X}}$.

The approximate MSE of the estimator $\bar{y}_{R1}$ can also be derived using the first order Taylor expansion. The MSE of $\bar{y}_{R1}$ is given by

$$MSE(\bar{y}_{R1}) = V(\bar{y}_M) + R^2 V(\bar{x}_M) - 2RCov(\bar{y}_M, \bar{x}_M). \quad (6)$$

In practice, the MSE will remain unknown but can be estimated from the sample data. For a sample estimate of the $MSE(\bar{y}_{R1})$, the sample estimate can be substituted as follows:

$$M\hat{S}E(\bar{y}_{R1}) = \hat{V}(\bar{y}_M) + \hat{R}^2 \hat{V}(\bar{x}_M) - 2\hat{R}C\hat{o}v(\bar{y}_M, \bar{x}_M), \quad (7)$$

where $\hat{R} = \frac{\bar{y}_M}{\bar{x}_M}$.

An alternative ratio estimator is derived from separate ratio estimators of $\bar{Y}_C$ and $\bar{Y}_{C'}$. Let $\bar{y}_{RC} = \frac{\bar{y}_C}{\bar{x}_C}\bar{X}_C$ and $\bar{y}_{RC'} = \frac{\bar{y}_{C'}}{\bar{x}_{C'}}\bar{X}_{C'}$ be the ratio estimators of $\bar{Y}_C$ and $\bar{Y}_{C'}$, respectively. These ratio estimates are weighted to provide an estimate of the population mean. The second ratio estimator of $\bar{Y}$ is given by

$$\bar{y}_{R2} = \hat{P}\frac{\bar{y}_C}{\bar{x}_C}\bar{X}_C + (1-\hat{P})\frac{\bar{y}_{C'}}{\bar{x}_C}\bar{X}_{C'}. \quad (8)$$

Information on $\bar{X}_C$ and $\bar{X}_{C'}$ is required for this ratio estimate.

Let $f_1 = \frac{m}{M}$ and $f_2 = \frac{n-m}{N-M}$ be the sampling fractions in classes *C* and *C'*, respectively. Meanwhile, let $B(\bar{y}_{RC})$ denote the bias of $\bar{y}_{RC}$ for estimating the parameter $\bar{Y}_C$. Similarly, $B(\bar{y}_{RC'})$ denotes the bias of $\bar{y}_{RC'}$ in $\bar{Y}_{C'}$ estimation. When the sample size is fixed, the approximate biases are given by

$$B(\bar{y}_{RC}) = \left(R_C\frac{S_{XC}^2}{\bar{X}_C} - \frac{S_{XYC}}{\bar{X}_C}\right)\frac{1}{m}(1-f_1), \text{ and}$$

$$B(\bar{y}_{RC'}) = \left(R_{C'}\frac{S_{XC'}^2}{\bar{X}_{C'}} - \frac{S_{XYC'}}{\bar{X}_{C'}}\right)\frac{1}{m-1}\hat{P}(1-f_2).$$

The bias of $\bar{y}_{R2}$ for estimating $\bar{Y}$ is equal to the following:

$$\begin{aligned}
B(\bar{y}_{R2}) &= E(\bar{y}_{R2}) - \bar{Y} \\
&= E\left[\hat{P}\frac{\bar{y}_C}{\bar{x}_C}\bar{X}_C + (1-\hat{P})\frac{\bar{y}_{C'}}{\bar{x}_C}\bar{X}_{C'}\right] - E\left[\hat{P}\bar{Y}_C + (1-\hat{P})\bar{Y}_{C'}\right] \\
&= E\left[\hat{P}\left(\frac{\bar{y}_C}{\bar{x}_C}\bar{X}_C - \bar{Y}_C\right) + (1-\hat{P})\left(\frac{\bar{y}_{C'}}{\bar{x}_C}\bar{X}_{C'} - \bar{Y}_{C'}\right)\right] \\
&= E\left[\hat{P}\left(\frac{\bar{y}_C}{\bar{x}_C}\bar{X}_C - \bar{Y}_C\right)\right] + E\left[(1-\hat{P})\left(\frac{\bar{y}_{C'}}{\bar{x}_C}\bar{X}_{C'} - \bar{Y}_{C'}\right)\right] \\
&= E_1\left[\hat{P}E_2\left(\frac{\bar{y}_C}{\bar{x}_C}\bar{X}_C - \bar{Y}_C\right)\right] + E_1\left[(1-\hat{P})E_2\left(\frac{\bar{y}_{C'}}{\bar{x}_C}\bar{X}_{C'} - \bar{Y}_{C'}\right)\right] \\
&= E_1\left[\hat{P}B(\bar{y}_{RC})\right] + E_1\left[(1-\hat{P})B(\bar{y}_{RC'})\right] \\
&= \left(R_C\frac{S_{XC}^2}{\bar{X}_C} - \frac{S_{XYC}}{\bar{X}_C}\right)\frac{1}{m}P(1-f_1) \\
&\quad + \left(R_{C'}\frac{S_{XC'}^2}{\bar{X}_{C'}} - \frac{S_{XYC'}}{\bar{X}_{C'}}\right)\frac{1}{m-1}E\left[\hat{P}(1-f_2)\right].
\end{aligned} \quad (9)$$

The bias of $\bar{y}_{R2}$ is combined, forming the biases of $\bar{y}_{RC}$ and $\bar{y}_{RC'}$. If $m \to \infty$, then $B(\bar{y}_{R2}) \to 0$.

Given the sample size, *n*, let V_1 and V_2 be the unconditional and conditional variances, respectively. As $m \to \infty$, the bias of $\bar{y}_{R2}$ is close to zero, that is, $E_2(\bar{y}_{RC} \mid n) = \bar{Y}_{RC}$ and $E_2(\bar{y}_{RC'} \mid n) = \bar{Y}_{RC'}$. The approximate MSE of $\bar{y}_{R2}$ is given by

$$\begin{aligned}
MSE(\bar{y}_{R2}) \approx V(\bar{y}_{R2}) &= E_1 V_2(\bar{y}_{R2}) + V_1 E_2(\bar{y}_{R2}) \\
&= E_1\left[\hat{P}^2\frac{(1-f_1)}{m}\left(S_{YC}^2 + R_C^2 S_{XC}^2 - 2R_C S_{XYC}\right)\right] \\
&\quad + E_1\left[\left(1-\hat{P}\right)^2\frac{(1-f_2)}{n-m}\left(S_{YC'}^2 + R_{C'}^2 S_{XC'}^2 - 2R_{C'} S_{XYC'}\right)\right] \\
&\quad + V_1\left\{\hat{P}\bar{Y}_C + \left(1-\hat{P}\right)\bar{Y}_{C'}\right\},
\end{aligned}$$

$$= E_1\left[\hat{P}^2\frac{(1-f_1)}{m}S_{RC}^2 + (1-\hat{P})^2\frac{(1-f_2)}{n-m}S_{RC'}^2\right]$$
$$+V_1\left[\hat{P}(\bar{Y}_C - \bar{Y}_{C'}) + \bar{Y}_{C'}\right],$$
$$= (\bar{Y}_C - \bar{Y}_{C'})^2 V(\hat{P}) + \frac{(1-f_1)S_{RC}^2}{m}E_1(\hat{P}^2)$$
$$+\frac{S_{RC'}^2}{m-1}E\left[\hat{P}(1-\hat{P})(1-f_2)\right], \quad (10)$$

where $S_{RC}^2 = S_{YC}^2 + R_C^2 S_{XC}^2 - 2R_C S_{XYC}$ and $S_{RC'}^2 = S_{YC'}^2 + R_{C'}^2 S_{XC'}^2 - 2R_{C'} S_{XYC'}$.

The sample estimates comprise $\hat{S}_{RC}^2 = \hat{S}_{YC}^2 + \hat{R}_C^2 \hat{S}_{XC}^2 - 2\hat{R}_C \hat{S}_{XYC}$ and $\hat{S}_{RC'}^2 = \hat{S}_{YC'}^2 + \hat{R}_{C'}^2 \hat{S}_{XC'}^2 - 2\hat{R}_{C'} \hat{S}_{XYC'}$, where $\hat{R}_C = \frac{\bar{y}_C}{\bar{x}_C}$ and $\hat{R}_{C'} = \frac{\bar{y}_{C'}}{\bar{x}_{C'}}$, to find the sample estimate of the $MSE(\bar{y}_{R2})$. Under inverse random sampling, $E_2(\bar{y}_{RC}^2 \mid n) = \bar{Y}_C^2 + \left(1-\frac{m}{M}\right)\frac{S_{RC}^2}{m}$, $E_2(\bar{y}_{RC'}^2 \mid n) = \bar{Y}_{C'}^2 + \left(1-\frac{n-m}{N-M}\right)\frac{S_{RC'}^2}{n-m}$. In addition, when the sample size was given, $E_2(\hat{S}_{RC}^2 \mid n) \approx S_{RC}^2$ and $E_2(\hat{S}_{RC'}^2 \mid n) \approx S_{RC'}^2$. For the sample estimate of the $MSE(\bar{y}_{R2})$, the sample estimate can be substituted as follows:

$$M\hat{S}E(\bar{y}_{R2}) = (\bar{y}_{RC} - \bar{y}_{RC'})^2 \hat{V}(\hat{P}) + \frac{\hat{S}_{RC}^2}{m}\left(\hat{P}^* - \frac{m}{N}\hat{P}\right)$$
$$+\hat{S}_{RC'}^2\left(\frac{\hat{P}(1-\hat{P})}{m-1} - \frac{1-\hat{P}}{N} - \frac{\hat{V}(\hat{P})}{n-m}\right). \quad (11)$$

3.2 Application to real data

To clearly illustrate the application with real data, the yield data of off-season rice in 2021 reported by Center of Agricultural Information, Office of Agricultural Economics was gathered from the publicly available official website https://www.oae.go.th and was used for the application. This dataset includes the yield of off-season rice and the cultivated area from only 71 provinces with off-season rice cultivation in Thailand in 2021. Hence, the population included 71 provinces in Thailand. The study variable (y) and the auxiliary variable (x) are the yield of off-season rice (ton) and the cultivated area (rai), respectively. A unit is defined in class C if the rice product of the unit is greater than or equal to 150,000 tons, that is, $C = \{u_i; y_i, \geq 150{,}000\}$ for i = 1, 2, … 71. The means of the cultivated area in classes C and C' are 403,117 and 58,273 rai, respectively. The mean of the overall cultivated area is 116,557 rai. Inverse random sampling with m = 3 is used to draw the units. Suppose that the sample comprises three units of interest at the 10-th draw. Table 1 represents the values of y and x.

As shown in Table 1, the sample comprising s = {1, 2, …, 10} is partitioned into two parts: part s_C = {4, 9, 10} and $s_{C'}$ = {1, 2, 3, 5, 6, 7, 8} with 3 and 7 units for each part respectively.

Table 2 represents the population mean estimate of the cultivated area and the yield of off-season rice in class $C; (\bar{x}_C), (\bar{y}_C)$ and $C'; (\bar{x}_{C'}), (\bar{y}_{C'})$, respectively. These values were used to calculate the unbiased estimate $(\bar{y}_M)$, the first ratio estimate $(\bar{y}_{R1})$, and the second ratio estimate $(\bar{y}_{R2})$, which are equal to 54,755.56, 70,182.48, and 78,975.98 tons, respectively.

Table 1. The values form the units under inverse random sample

i-th draw	Cultivated area (x)	Rice product (y)
1	500	193
2	19,233	11,440
3	23,582	10,189
4	450,946	285,077
5	2,433	1,535
6	17,160	10,084
7	40,494	23,460
8	16,902	8,295
9	343,078	195,499
10	288,160	180,894

Table 2. Sample mean of the cultivated area and the rice product in class C and C'

Statistics	S_{C}		$S_{C'}$	
	x (rai)	y(ton)	x(rai)	y(ton)
Number of units	m = 3		n-m = 7	
Sample mean	360,728	220,490	13,853	7,403

3.3 Simulation study

The simulation study comprises 50,000 replications. The population mean ($\bar{Y}$) was estimated for each sample. The values of the estimates $(\bar{y})$ and the final sample size (n) are averaged in each simulation time. The averages were interpreted as expected values, that is,

$$\tilde{E}(\bar{y}) = \frac{1}{50{,}000}\sum_{i=1}^{50{,}000}\bar{y}_i \text{ and } \tilde{E}(n) = \frac{1}{50{,}000}\sum_{i=1}^{50{,}000}n_i .$$

The MSE estimate is calculated as follows:

$$M\tilde{S}E(\bar{y}) = \frac{1}{50{,}000}\sum_{i=1}^{50{,}000}(\bar{y}_i - \bar{Y})^2 .$$

The bias of an estimator is given by

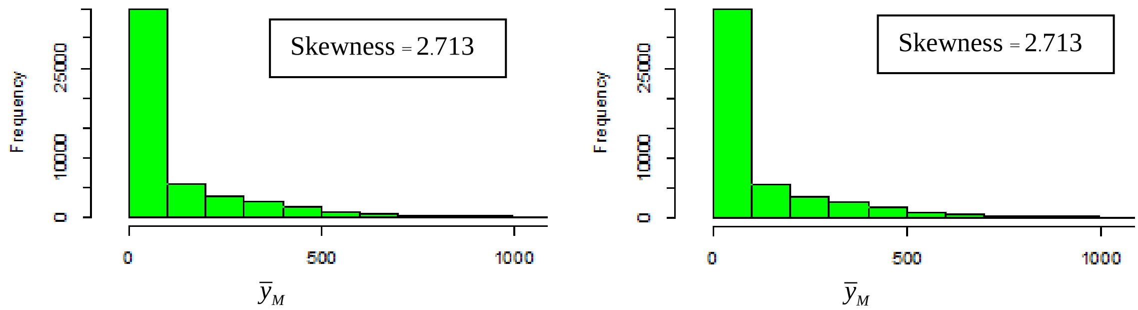
$$\tilde{B}(\bar{y}) = \tilde{E}(\bar{y}) - \bar{Y} .$$

The relative efficiency of the estimator is calculated as follows:

$$eff(\bar{y}) = \frac{M\tilde{S}E(\bar{y}_M)}{M\tilde{S}E(\bar{y})},$$

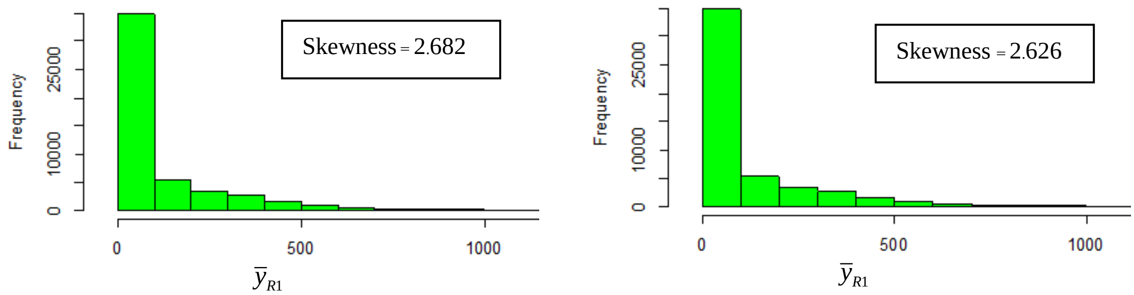
where $\bar{y}_M$ is the unbiased estimator, and $\bar{y}$ stands for $\bar{y}_{R1}$ and $\bar{y}_{R2}$.

For the case m = 5, the estimates from the three estimators were used to create histogram plots, as shown in Figures 1–3. These figures represent the histogram of the



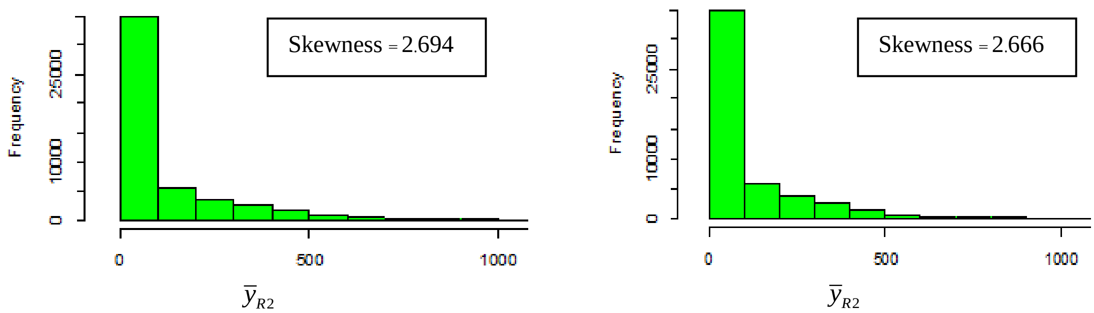
a) Correlation coefficient = 0.3   b) Correlation coefficient = 0.7

Figure 1. The histrogram of $\bar{y}_M$ with m = 5 (a) with correlation coefficient 0.3 between y and x, and (b) with correlation coefficient 0.7 between y and x



a) Correlation coefficient = 0.3   b) Correlation coefficient = 0.7

Figure 2. The histrogram of $\bar{y}_{R1}$ with m = 5 (a) with correlation coefficient 0.3 between y and x, and (b) with correlation coefficient 0.7 between y and x



a) Correlation coefficient = 0.3   b) Correlation coeffieint = 0.7

Figure 3. The histrogram of $\bar{y}_{R2}$ with m = 5 (a) with correlation coefficient 0.3 between y and x, and (b) with correlation coefficient 0.7 between y and x

estimators $\bar{y}_M, \bar{y}_{R1}, \bar{y}_{R2}$ with skewness coefficient. The findings indicate that the distributions of all estimators were right-skewed. From Figure 1, when the correlation coefficient increases it can be observed that the skewness coefficient of the estimator $\bar{y}_M$ remains constant. This is attributed to the fact that the estimator $\bar{y}_M$ does not incorporate the auxiliary variable in the estimation process. From Figures 2 and 3, it can be observed that as the correlation coefficient increases, the skewness coefficient decreases. When the correlation coefficient is held constant, the estimator $\bar{y}_{R1}$ exhibits the lowest skewness, followed by the estimator $\bar{y}_{R2}$, and subsequently $\bar{y}_M$.

Table 3 demonstrates an increase in m as the average sample size rises. The average values of the unbiased estimator are close to the true population mean for all situations. As a result, the estimated biases of the unbiased estimator are close to zero. The results are consistent with the property of an unbiased estimator. Given the coefficient of correlation and the number of units of interest in the sample, the biased sizes of the estimator $\bar{y}_{R2}$ are larger than those of the estimator $\bar{y}_{R1}$.

The biased sizes of two ratio estimators are small when the correlation coefficient between auxiliary and study variables is equal to 0.1. Given the number of units of interest in the sample, m, the biased size of the two ratio estimators increases with the correlation coefficient. For $\rho \geq 0.7$, the biased size of the estimator $\bar{y}_{R2}$ decreases as the number of units of interest in the sample increases.

Table 4 shows that the MSE of the estimator $\bar{y}_{R2}$ is the lowest in all scenarios. The MSEs of estimator $\bar{y}_{R1}$ are lower than those of the unbiased estimator $\bar{y}_{M}$. When the correlation coefficient was given, the MSE of the three estimators decreased as the number of units of interest in the sample increased. Similarly, given the number of units of interest in the sample, the MSE of both ratio estimators decreases when the correlation coefficient increases.

The relative efficacies of estimators $\bar{y}_{R1}$ and $\bar{y}_{R2}$ are greater than 1. Therefore, both ratio estimators yield higher efficiency compared with the unbiased estimator. When the correlation coefficient is provided, the relative efficiencies of the two ratio estimators decrease as the number of units of interest in the sample increases. When the number of units of interest in the sample is fixed, an increase in the correlation coefficient leads to a rise in the relative efficiencies of both ratio estimators.

Table 3. The average values of the estimates and the estimate biases of the estimator

ρ	m	$\tilde{E}(n)$	$\tilde{E}(\bar{y}_M)$	$\tilde{E}(\bar{y}_{R1})$	$\tilde{E}(\bar{y}_{R2})$	$\tilde{B}(\bar{y}_M)$	$\tilde{B}(\bar{y}_{R1})$	$\tilde{B}(\bar{y}_{R2})$
0.1	3	21.50	116.11	115.94	115.19	-0.556	-0.727	-1.474
	4	28.69	117.05	116.94	116.37	0.387	0.275	-0.294
	5	35.80	117.42	117.34	116.93	0.756	0.676	0.261
	6	43.03	116.62	116.57	116.23	-0.042	-0.097	-0.431
	7	50.26	116.19	116.15	115.86	-0.478	-0.518	-0.806
0.3	3	21.50	116.11	115.51	113.35	-0.556	-1.151	-3.312
	4	28.69	117.05	116.66	115.04	0.387	-0.004	-1.629
	5	35.80	117.42	117.14	115.89	0.756	0.478	-0.777
	6	43.03	116.62	116.42	115.42	-0.042	-0.245	-1.249
	7	50.26	116.19	116.03	115.20	-0.478	-0.633	-1.469
0.5	3	21.50	116.11	115.01	111.27	-0.556	-1.654	-5.391
	4	28.69	117.05	116.33	113.50	0.387	-0.338	-3.161
	5	35.80	117.42	116.90	114.69	0.756	0.239	-1.978
	6	43.03	116.62	116.24	114.46	-0.042	-0.422	-2.201
	7	50.26	116.19	115.89	114.42	-0.478	-0.770	-2.244
0.7	3	21.50	116.11	114.21	108.17	-0.556	-2.458	-8.495
	4	28.69	117.05	115.79	111.17	0.387	-0.876	-5.490
	5	35.80	117.42	116.52	112.84	0.756	-0.149	-3.826
	6	43.03	116.62	115.95	112.99	-0.042	-0.712	-3.677
	7	50.26	116.19	115.67	113.21	-0.478	-0.995	-3.452
0.9	3	21.50	116.11	112.12	101.17	-0.556	-4.548	-15.491
	4	28.69	117.05	114.36	105.72	0.387	-2.306	-10.941
	5	35.80	117.42	115.47	108.41	0.756	-1.191	-8.252
	6	43.03	116.62	115.17	109.40	-0.042	-1.496	-7.268
	7	50.26	116.19	115.06	110.24	-0.478	-1.606	-6.426

Table 4. The estimates of MSE and relative efficacy of estimators

ρ	m	$\tilde{MSE}(\bar{y}_M)$	$\tilde{MSE}(\bar{y}_{R1})$	$\tilde{MSE}(\bar{y}_{R2})$	$eff(\bar{y}_{R1})$	$eff(\bar{y}_{R2})$
0.1	3	62014.97	61519.21	60382.08	1.008	1.027
	4	40489.32	40278.28	39665.31	1.005	1.021
	5	28701.78	28591.80	28257.94	1.004	1.016
	6	21359.83	21308.39	21092.33	1.002	1.013
	7	16534.05	16506.95	16348.24	1.002	1.011
0.3	3	62014.97	60399.85	57366.86	1.027	1.081
	4	40489.32	39808.48	38186.90	1.017	1.060
	5	28701.78	28365.93	27409.68	1.012	1.047
	6	21359.83	21181.35	20565.52	1.008	1.039
	7	16534.05	16430.53	15997.90	1.006	1.034
0.5	3	62014.97	59101.33	54066.92	1.049	1.147
	4	40489.32	39254.93	36530.81	1.031	1.108
	5	28701.78	28097.06	26446.49	1.022	1.085
	6	21359.83	21029.44	19961.79	1.016	1.070
	7	16534.05	16338.80	15593.69	1.012	1.060
0.7	3	62014.97	57094.99	49356.94	1.086	1.256
	4	40489.32	38381.19	34093.07	1.055	1.188
	5	28701.78	27666.65	25002.49	1.037	1.148
	6	21359.83	20784.64	19045.18	1.028	1.122
	7	16534.05	16190.18	14974.34	1.021	1.104

Table 4. Continued.

ρ	m	$M\tilde{S}E(\bar{y}_M)$	$M\tilde{S}E(\bar{y}_{R1})$	$M\tilde{S}E(\bar{y}_{R2})$	$eff(\bar{y}_{R1})$	$eff(\bar{y}_{R2})$
0.9	3	62014.97	52233.39	39663.27	1.187	1.564
	4	40489.32	36167.40	28764.69	1.119	1.408
	5	28701.78	26542.61	21727.36	1.081	1.321
	6	21359.83	20135.99	16911.32	1.061	1.263
	7	16534.05	15791.66	13504.29	1.047	1.224

4. Conclusions

Two ratio estimators of the population mean were developed in this study under inverse random sampling. The first ratio estimator utilizes the information from the values of the auxiliary variable in the sample combined with the entire population mean of the auxiliary variable. The second ratio estimator utilizes the information from the values of the auxiliary variable in the sample with the population mean in each class of population. In the simulation study, the second ratio estimator yields higher biases than those of the first ratio estimator. This result may be attributed to the bias of the second ratio estimator is associated with the biases of two ratio estimators of the population mean in each class. The developed ratio estimators yield a smaller MSE than the unbiased estimator. The second ratio estimator reveals the highest relative efficiency, especially when the correlation coefficient is high. This finding is probably attributable to the use of additional information by the second ratio estimator. The results of this study can be used when employing auxiliary variables to improve parameter estimates for rare populations.

Acknowledgements

This work was supported by the Faculty of Science, Silpakorn University, Nakorn Pathom, grant number SRIFJRG-2564-13.

References

Center of Agricultural Information, Office of Agricultural Economics. (2021). Retrieved from https://www.oae.go.th

Christman, M. C., & Lan, F. (2001). Inverse adaptive cluster sampling. *Biometrics*, *57*(4), 1096-1105. Retrieved from https://www.jstor.org/stable/3068241

Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). New York, NY: John Wiley and Sons.

Finney, D. J. (1949). On a method of estimating frequencies. *Biometrika*, *36*(1), 223–234. doi:10.1093/biomet/36.1-2.233

Greco, L., & Naddeo, S. (2007). Inverse sampling with unequal selection probabilities. *Communications in Statistics: Theory and Methods, 36*(5), 1039-1048. doi:10.1080/03610920601033926

Haldane, J. B. S. (1945). On a method of estimating frequencies. *Biometrika, 33*, 222–225. doi:10.2307/2332299

Murthy, M. N. (1964). Product method of estimation. *Sankhya, 26*, 69-74. Retrieved from https://www.jstor.org/stable/25049308

Salehi, M. M., & Seber, G. A. F. (2001). A new proof of Murthy's estimator with applies to sequential sampling. *Australian and New Zealand Journal of Statistics*, *43*(3), 281-286. doi:10.1111/1467-842X.00174

Sangngam, P. (2014). Ratio estimators using coefficient of variation and coefficient of correlation. *Modern Applied Science*, *8*(5), 70-79. doi:10.5539/mas.v8n5p70

Sangngam, P., & Suwattee, P. (2012). Stratified inverse sampling for rare populations. *Thailand Statistician, 10*(1), 69–86. Retrieved from https://ph02.tci-thaijo.org/index.php/thaistat/article/view/34233

Sharma, V., & Kumar, S. (2021). Class of ratio-cum-product type estimator under double sampling: A simulation study. *Thailand Statistician*, *19*(4), 734–742. Retrieved from https://ph02.tci-thaijo.org/index.php/thaistat/article/view/245234

Smith, D. R, Conroy, M. J., & Bralhage, D. H. (1995). Efficiency of adaptive cluster sampling for estimating density of wintering waterfowl. *Biometrics*, *51*, 777-788. doi:10.2307/2532964

Sungsuwan, S., & Suwattee, P. (2014). Model-assisted estimation in inverse sampling. *Chiang Mai Journal of Science*, *41*(3), 704-713. Retrieved from https://epg.science.cmu.ac.th/ejournal/journal-detail.php?id=5037