Modern Applied Science; Vol. 8, No. 5; 2014
ISSN 1913-1844 E-ISSN 1913-1852
Published by Canadian Center of Science and Education

Ratio Estimators Using Coefficient of Variation and Coefficient of Correlation

Prayad Sangngam[1]

[1] Department of Statistics, Faculty of Science, Silpakorn University, Nakorn Pathom, Thailand

Correspondence: Prayad Sangngam, Department of Statistics, Faculty of Science, Silpakorn University, Nakorn Pathom 73000, Thailand. E-mail: prayad@su.ac.th

Received: June 4, 2014 Accepted: June 10, 2014 Online Published: August 6, 2014

doi:10.5539/mas.v8n5p70 URL: http://dx.doi.org/10.5539/mas.v8n5p70

Abstract

This paper introduces ratio estimators of the population mean using the coefficient of variation of study variable and auxiliary variables together with the coefficient of correlation between the study and auxiliary variables under simple random sampling and stratified random sampling. These ratio estimators are almost unbiased. The mean square errors of the estimators and their estimators are given. Sample size estimation in both sampling designs are presented. An optimal sample size allocation in stratified random sampling is also suggested. Based on theoretical study, it can be shown that these ratio estimators have smaller MSE than the unbiased estimators. Moreover, the empirical study indicates that these ratio estimators have smallest MSE compared to the existing ones.

Keywords: ratio estimator, sample size estimation, coefficient of variation, coefficient of correlation

1. Introduction

Consider a population of N units with observations (x_i, y_i) for $i = 1, 2, \ldots, N$ where y_i is a value of study variable and x_i is a value of auxiliary variable. Under simple random sampling without replacement, an unbiased estimator of the population mean $\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i$ is the sample mean $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$. The variance of the unbiased estimator is

$$V(\bar{y}) = \frac{1-f}{n} S_y^2, \tag{1.1}$$

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

A common 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 auxiliary variable (C_x) and coefficient of variation of study variable (C_y). Murthy (1964) has suggested that if $\rho > \frac{C_x}{2C_y}$, the ratio estimator performs better than 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:

$$B(\bar{y}_R) = \frac{1-f}{n}\left(\frac{R}{\bar{X}} S_x^2 - \frac{1}{\bar{X}} S_{xy}\right), \tag{1.2}$$

$$MSE(\bar{y}_R) = \frac{1-f}{n}\left(S_y^2 + R^2 S_x^2 - 2RS_{xy}\right), \tag{1.3}$$

where $R = \frac{\bar{Y}}{\bar{X}}$, $S_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{X})^2$ and $S_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{Y})(x_i - \bar{X})$. When the C_x is known, Sissodia and Dwivedi (1981) has proposed a modified ratio estimator, $\bar{y}_{SD} = \bar{y}\dfrac{\bar{X}+C_x}{\bar{x}+C_x}$. The approximate bias and MSE of the estimator are

$$B(\bar{y}_{SD}) = \frac{1-f}{n}\left(\frac{R_x}{\bar{X}}S_x^2 - \frac{1}{\bar{X}}S_{xy}\right), \tag{1.4}$$

$$MSE(\bar{y}_{SD}) = \frac{1-f}{n}\left(S_y^2 + R_x^2S_x^2 - 2R_xS_{xy}\right), \tag{1.5}$$

where $R_x = \dfrac{\bar{Y}}{\bar{X}+C_x}$. Sampath (2005) used the coefficient of variation of the study variable to improve the unbiased estimator as $\bar{y}_S = \left(1+\dfrac{1-f}{n}C_y^2\right)^{-1}\bar{y}$. The approximate bias and MSE of this estimator are

$$B(\bar{y}_S) = \left(1+\frac{1-f}{n}C_y^2\right)^{-1}\bar{Y} - \bar{Y}, \tag{1.6}$$

$$MSE(\bar{y}_S) = \frac{1-f}{n}S_y^2\left(1+\frac{1-f}{n}C_y^2\right)^{-1}. \tag{1.7}$$

In addition, there are several authors, such as Upadhyaya and Singh (1999), Singh and Tailor (2003), who have developed various ratio estimators under simple random sampling.

If the study variable has different mean values in different subpopulations, it is advantageous to draw a sample by stratified random sampling. In stratified sampling, a population is partitioned into L strata. A stratum h contains N_h units with observations (x_{hi}, y_{hi}) where $h = 1, 2, \ldots, L$ and $i = 1, 2, \ldots, N_h$. An unbiased estimator of $\bar{Y}$ under stratified random sampling is given by $\bar{y}_{st} = \sum_{h=1}^{L} W_h\bar{y}_h$ where $W_h = \dfrac{N_h}{N}$ is the stratum weight and $\bar{y}_h$ is the sample mean of the study variable in stratum h. The variance of the unbiased estimator is

$$V(\bar{y}_{st}) = \sum_{h=1}^{L} W_h^2\gamma_h S_{yh}^2, \tag{1.8}$$

where $\gamma_h = (1-f_h)/n_h$, $f_h = \dfrac{n_h}{N_h}$ is sampling fraction in stratum h, n_h is a sample size in stratum h and S_{yh}^2 is the variance of the study variable in stratum h. There are two types of ratio estimators in stratified random sampling, namely combined and separate ratio estimators.

The combined ratio estimator is given by $\bar{y}_{RC} = \dfrac{\bar{y}_{st}}{\bar{x}_{st}}\bar{X}$, where $\bar{x}_{st} = \sum_{h=1}^{L} W_h\bar{x}_h$ is an unbiased estimator of the population mean $\bar{X}$ and $\bar{x}_h$ is the sample mean of auxiliary variable in stratum h (Cochran, 1977). The approximate mean squared error of the combined ratio estimator is

$$MSE(\bar{y}_{RC}) = \sum_{h=1}^{L} W_h^2\gamma_h\left(S_{yh}^2 + R^2S_{xh}^2 - 2RS_{xyh}\right), \tag{1.9}$$

where $R = \dfrac{\bar{Y}}{\bar{X}}$ is the population ratio, S_{xh}^2 is the variance of auxiliary variable in stratum h and S_{xyh} is the covariance between auxiliary and study variables in stratum h. The approximate bias of the combined ratio estimator is

$$B(\bar{y}_{RC}) = \sum_{h=1}^{L} W_h^2\gamma_h\left(\frac{R}{\bar{X}}S_{xh}^2 - \frac{1}{\bar{X}}S_{xyh}\right). \tag{1.10}$$

The separate ratio estimator is given by $\bar{y}_{RS} = \sum_{h=1}^{L} W_h\dfrac{\bar{y}_h\bar{X}_h}{\bar{x}_h}$. The approximate MSE of the separate ratio estimator can be given by

$$MSE(\bar{y}_{RS}) = \sum_{h=1}^{L} W_h^2\gamma_h\left(S_{yh}^2 + R_h^2S_{xh}^2 - 2R_hS_{xyh}\right), \tag{1.11}$$

where $R_h = \frac{\bar{Y}_h}{\bar{X}_h}$ is the population ratio in stratum h. The approximate bias of the separate ratio estimator is

$$B(\bar{y}_{RS}) = \sum_{h=1}^{L} W_h \gamma_h \left(\frac{R_h}{\bar{X}_h} S_{xh}^2 - \frac{1}{\bar{X}_h} S_{xyh} \right). \tag{1.12}$$

Kadilar and Cingi (2003) have proposed several combined ratio estimators. The simplest one based on the Sissodia and Dwivedi (1981) estimator is defined as $\bar{y}_{KC} = \bar{y}_{st} \frac{\sum_{h=1}^{L} W_h (\bar{X}_h + C_{xh})}{\sum_{h=1}^{L} W_h (\bar{x}_h + C_{xh})}$ where C_{xh} is the coefficient of variation of the auxiliary variable in stratum h. The MSE and bias of this estimator are approximated as follows:

$$MSE(\bar{y}_{KC}) = \sum_{h=1}^{L} W_h^2 \gamma_h \left(S_{yh}^2 + R_{KC}^2 S_{xh}^2 - 2R_{KC} S_{xyh} \right), \tag{1.13}$$

$$B(\bar{y}_{KC}) = \sum_{h=1}^{L} W_h^2 \gamma_h \left(\frac{R_{KC}}{\bar{X}_{KC}} S_{xh}^2 - \frac{1}{\bar{X}_{KC}} S_{xyh} \right), \tag{1.14}$$

where $R_{KC} = \frac{\bar{Y}_{st}}{\bar{X}_{KC}} = \frac{\bar{Y}_{st}}{\sum_{h=1}^{L} W_h (\bar{X}_h + C_{xh})}$.

Kadilar and Cingi (2005) have improved the combined ratio estimator in stratified random sampling based on the estimator introduced by Prasad (1989). However, these estimators depend on several unknown parameters and therefore aredifficult to use.

In Sections 2 and 3, the ratio estimators based on the coefficient of variation and correlation in simple random sampling and stratified random sampling are introduced, respectively. The approximate bias and MSE of the estimator are derived. An estimator of the MSE is given. The sample size estimation and an optimal allocation of sample size in stratified random sampling is presented. The comparison of the efficiency between the proposed estimator and unbiased estimator is theoretically provided. Hypothetical populations are used to compare the properties of the presented estimators with the existing ones.

2. Estimation in Simple Random Sampling

2.1 Parameter Estimation

Consider the following ratio estimator for the population mean of the study variable,

$$\bar{y}_c = \bar{y} \frac{\bar{X} + c}{\bar{x} + c} \tag{2.1}$$

where c is a real constant to be determined such that the $MSE(\bar{y}_c)$ is minimized. Note that when c is equal to 0, this estimator is reduced to the usual ratio estimator and when c is equal to C_x this estimator become the estimator of Sissodia and Dwivedi (1981). To obtain the MSE and bias of the estimator (2.1), let $e_1 = \frac{\bar{y} - \bar{Y}}{\bar{Y}}$ and $e_2 = \frac{\bar{x} - \bar{X}}{\bar{X} + c}$. It can be shown that $E(e_1) = 0$, $E(e_2) = 0$, $E(e_1^2) = \frac{V(\bar{y})}{\bar{Y}^2}$, $E(e_1 e_2) = \frac{Cov(\bar{x}, \bar{y})}{\bar{Y}(\bar{X} + c)}$ and $E(e_2^2) = \frac{V(\bar{x})}{(\bar{X} + c)^2}$. The estimator $\bar{y}_c$ can be written as $\bar{y}_c = \bar{Y}(1 + e_1)(1 + e_2)^{-1}$. Using Taylor series approximation, we obtain $\bar{y}_c = \bar{Y}(1 + e_1 - e_2 + e_2^2 - e_1 e_2 + ...)$. When the terms of degree greater than two are ignored, we get the approximate bias of the estimator $\bar{y}_c$ as

$$B(\bar{y}_c) = \left(\frac{1-f}{n} \right) \left[\frac{R_c S_x^2}{(\bar{X} + c)} - \frac{S_{xy}}{(\bar{X} + c)} \right], \tag{2.2}$$

where $R_c = \frac{\bar{Y}}{\bar{X}+c}$. Similarly, Taylor's formula can be used to approximate MSE of the estimator as

$$MSE(\bar{y}_c) = \frac{1-f}{n}\left(S_y^2 + R_c^2 S_x^2 - 2R_c S_{xy}\right). \quad (2.3)$$

Minimizing (2.3) with respect to c, we get the optimum value of c as $c = c^* = \frac{\bar{Y}S_x^2}{S_{xy}} - \bar{X}$. Substituting c^* for c in (2.1), (2.2) and (2.3) and using algebra, we obtain the optimum estimator, its bias and MSE as follows,

$$\bar{y}_{c^*} = \bar{y}\frac{\bar{X}C_x}{C_y\rho(\bar{x}-\bar{X}) + \bar{X}C_x}, \quad (2.4)$$

$$B(\bar{y}_{c^*}) = 0, \quad (2.5)$$

$$MSE(\bar{y}_{c^*}) = \frac{(1-f)}{n}S_y^2(1-\rho^2). \quad (2.6)$$

Note that the optimum estimator is almost unbiased and its MSE is always smaller than the variance of the unbiased estimator. In addition, the optimum estimator can be applied for both populations with positive and negative coefficient of correlation. For a sample estimate of the MSE, one can substitute the sample estimate of S_y^2 which gives

$$\hat{MSE}(\bar{y}_{c^*}) = \frac{(1-f)}{n}s_y^2(1-\rho^2), \quad (2.7)$$

where s_y^2 is the sample variance of the study variable.

2.2 Sample Size Estimation

Sample size estimation is one of the important aspects in sample surveys. If the sample size is too small, the sampling error may be too large. However, too large sample size implies a waste of resources. We would like to specify a sample size that is sufficiently large to ensure a high probability that the estimate closes to the parameter. Under simple random sampling, the population mean of the study variable $(\bar{Y})$ is estimated with the optimum estimator $\bar{y}_{c*}$. To obtain the desired sample size, one can specify the margin of error d and the probability α such that $P\left(\left|\bar{y}_{c*} - \bar{Y}\right| \geq d\right) = \alpha$. Under some technical conditions as shown in Scott and Wu (1981) and Hajek (1960), we can show that $\bar{y}_{c*}$ is asymptotically normal distributed with mean $\bar{Y}$ and variance $MSE(\bar{y}_{c*})$. To obtain the absolute precision, we can find a value of n that satisfies $d / \sqrt{MSE(\bar{y}_{c*})} = z_{\alpha/2}$ where $z_{\alpha/2}$ denotes the upper $\alpha/2$ point of the standard normal distribution. Solving for n, we have

$$n = \frac{n_0}{1+\frac{n_0}{N}}, \quad (2.8)$$

where $n_0 = \frac{z_{\frac{\alpha}{2}}^2 S_y^2(1-\rho^2)}{d^2}$. If the population size N is large relative to the sample size n, the formula of the sample size reduces to n_0.

2.3 Comparison of Efficiency

In this section, the properties of the estimators in simple random sampling are compared. The relative efficiency of the optimum estimator and unbiased estimator is considered as follows:

$$e(\bar{y}, \bar{y}_{c*}) = \frac{E(\bar{y}-\bar{Y})^2}{E(\bar{y}_{c*}-\bar{Y})^2} = \frac{1}{1-\rho^2}. \quad (2.9)$$

This shows that the optimum estimator is always more efficient than the unbiased estimator because $0 < \rho^2 < 1$. The efficiency depends on the coefficient of correlation:if the coefficient of correlation increases then the efficiency also increases.

To compare the properties of the optimum estimator with the others, we consider hypothetical populations with vary characteristics. In this work, the coefficients of correlation in the populations are $\rho = 0.1, 0.2, \ldots, 0.9$. In each population, the coefficients of variations are $C_x = 0.2$, $C_y = 0.2$ and the population means $\bar{X} = 5,000$, $\bar{Y} = 5,000$. With varying sample sizes, the biases and MSEs of the estimators are given in Table 1 and Table 2, respectively. The biases and MSEs are computed by the formulas in the previous sections.

In Table 1, as expected, the absolute bias of the unbiased and optimum ratio estimators are always equal to 0. The estimator $\bar{y}_S$ has the largest absolute bias among the compared estimators. The bias of the estimator $\bar{y}_S$ is negative because the estimator is constructed by using a constant in which its value less than 1 multiplying the unbiased estimator. The bias of the estimator $\bar{y}_S$ does not depend on the coefficient of correlation. Observe that the absolute biases of the estimator $\bar{y}_{SD}$ are smaller than of the estimator $\bar{y}_R$. Given a sample size, when the coefficient of correlation increases the absolute bias of the two estimators $\bar{y}_R$ and $\bar{y}_{SD}$ decrease. Given a coefficient of correlation, the absolute bias of $\bar{y}_S$, $\bar{y}_R$ and $\bar{y}_{SD}$ decrease when the sample size increases.

Table 2 shows that the optimum ratio estimator has smallest MSE among the compared estimators. The MSEs of the two estimators $\bar{y}$ and $\bar{y}_S$ do not depend on the coefficient of correlation. When $\rho > 0.5$ the MSEs of the two estimators $\bar{y}_R$ and $\bar{y}_{SD}$ are less than those of the unbiased estimator. Given a sample size, when the coefficient of correlation increases the MSEs of the three estimators $\bar{y}_R$, $\bar{y}_{SD}$ and $\bar{y}_{c*}$ decrease. Given a coefficient of correlation, the MSEs of all estimators decrease when the sample size increases.

Table 1. Biases of the estimators in simple random sampling

n	ρ	$B(\bar{y})$	$B(\bar{y}_R)$	$B(\bar{y}_{SD})$	$B(\bar{y}_S)$	$B(\bar{y}_{c*})$
	0.1	0	5.9964	5.9961	-6.6538	0
	0.2	0	5.3301	5.3299	-6.6538	0
	0.3	0	4.6639	4.6636	-6.6538	0
	0.4	0	3.9976	3.9973	-6.6538	0
30	0.5	0	3.3313	3.3311	-6.6538	0
	0.6	0	2.6651	2.6648	-6.6538	0
	0.7	0	1.9988	1.9985	-6.6538	0
	0.8	0	1.3325	1.3323	-6.6538	0
	0.9	0	0.6663	0.6660	-6.6538	0
	0.1	0	3.5964	3.5962	-3.9928	0
	0.2	0	3.1968	3.1966	-3.9928	0
	0.3	0	2.7972	2.7970	-3.9928	0
	0.4	0	2.3976	2.3974	-3.9928	0
50	0.5	0	1.9980	1.9978	-3.9928	0
	0.6	0	1.5984	1.5982	-3.9928	0
	0.7	0	1.1988	1.1986	-3.9928	0
	0.8	0	0.7992	0.7990	-3.9928	0
	0.9	0	0.3996	0.3994	-3.9928	0
	0.1	0	1.7964	1.7963	-1.9952	0
	0.2	0	1.5968	1.5967	-1.9952	0
	0.3	0	1.3972	1.3971	-1.9952	0
	0.4	0	1.1976	1.1975	-1.9952	0
100	0.5	0	0.9980	0.9979	-1.9952	0
	0.6	0	0.7984	0.7983	-1.9952	0
	0.7	0	0.5988	0.5987	-1.9952	0
	0.8	0	0.3992	0.3991	-1.9952	0
	0.9	0	0.1996	0.1995	-1.9952	0

Table 2. MSEs of the estimators in simple random sampling

n	ρ	$MSE(\bar{y})$	$MSE(\bar{y}_R)$	$MSE(\bar{y}_{SD})$	$MSE(\bar{y}_S)$	$MSE(\bar{y}_{c*})$
	0.1	33313.33	59964.00	59961.60	33269.00	32980.20
	0.2	33313.33	53301.33	53299.20	33269.00	31980.80
	0.3	33313.33	46638.67	46636.80	33269.00	30315.13
	0.4	33313.33	39976.00	39974.40	33269.00	27983.20
30	0.5	33313.33	33313.33	33312.00	33269.00	24985.00
	0.6	33313.33	26650.67	26649.60	33269.00	21320.53
	0.7	33313.33	19988.00	19987.20	33269.00	16989.80
	0.8	33313.33	13325.33	13324.80	33269.00	11992.80
	0.9	33313.33	6662.67	6662.40	33269.00	6329.53
	0.1	19980.00	35964.00	35962.56	19964.04	19780.20
	0.2	19980.00	31968.00	31966.72	19964.04	19180.80
	0.3	19980.00	27972.00	27970.88	19964.04	18181.80
	0.4	19980.00	23976.00	23975.04	19964.04	16783.20
50	0.5	19980.00	19980.00	19979.20	19964.04	14985.00
	0.6	19980.00	15984.00	15983.36	19964.04	12787.20
	0.7	19980.00	11988.00	11987.52	19964.04	10189.80
	0.8	19980.00	7992.00	7991.68	19964.04	7192.80
	0.9	19980.00	3996.00	3995.84	19964.04	3796.20
	0.1	9980.00	17964.00	17963.28	9976.02	9880.20
	0.2	9980.00	15968.00	15967.36	9976.02	9580.80
	0.3	9980.00	13972.00	13971.44	9976.02	9081.80
	0.4	9980.00	11976.00	11975.52	9976.02	8383.20
100	0.5	9980.00	9980.00	9979.60	9976.02	7485.00
	0.6	9980.00	7984.00	7983.68	9976.02	6387.20
	0.7	9980.00	5988.00	5987.76	9976.02	5089.80
	0.8	9980.00	3992.00	3991.84	9976.02	3592.80
	0.9	9980.00	1996.00	1995.92	9976.02	1896.20

3. Estimation in Stratified Random Sampling

3.1 Parameter Estimation

In stratified random sampling, when $\bar{X}_h$, C_{xh} , C_{yh} and ρ_{xh} in stratum h are known, the separate ratio estimator can be modified as

$$\bar{y}_{RS\_C} = \sum_{h=1}^{L} W_h \frac{\bar{y}_h \bar{X}_h C_{xh}}{C_{yh}\rho_h\left(\bar{x}_h - \bar{X}_h\right) + \bar{X}_h C_{xh}} . \quad (3.1)$$

Since this estimator is constructed from the optimum ratio estimator, we call this estimator "optimum separate ratio estimator". To obtain the MSE and bias of the optimum separate ratio estimator, applying the MSE and bias of $\bar{y}_{c*h} = \frac{\bar{y}_h\left(\bar{X}_h + C_{xh}\right)}{\bar{x}_h + C_{xh}}$ under simple random sampling to draw in stratum h,yields

$$B\left(\bar{y}_{RS\_C}\right) = 0 , \quad (3.2)$$

$$MSE\left(\bar{y}_{RS\_C}\right) = \sum_{h=1}^{L} W_h^2 \gamma_h S_{yh}^2 \left(1 - \rho_h^2\right) , \quad (3.3)$$

For estimating the $MSE\left(\bar{y}_{RS\_C}\right)$, we substitute the sample estimates to obtain

$$\hat{MSE}\left(\bar{y}_{RS\_C}\right)=\sum_{h=1}^{L}W_h^2\gamma_h s_{yh}^2\left(1-\rho_h^2\right). \tag{3.4}$$

Note that the bias of the optimum separate ratio estimator is the cumulative bias of an optimum ratio estimate in each stratum which closes to zero. In addition, we found that the MSE of the estimator is also smaller than the variance of the unbiased estimator in (1.8).

3.2 Optimum Sample Size Allocation

Given a total sample size n and using the optimum separate ratio estimator, one may choose how to allocate the sample size among the L strata. In this section, the allocation scheme which minimizes the MSE of the estimator by fixing the total sample size is considered. That is, we need the values of $n_1, n_2, \ldots, n_L$ which minimize $MSE\left(\bar{y}_{RS\_C}\right)=\sum_{h=1}^{L}W_h^2\gamma_h S_{yh}^2\left(1-\rho_h^2\right)$ subject to the condition $n=\sum_{h=1}^{L}n_h$. The sample size allocated to each stratum is

$$n_h = n\frac{N_h S_h\sqrt{1-\rho_h^2}}{\sum_{h=1}^{L}N_h S_h\sqrt{1-\rho_h^2}} \qquad ; h=1,2,\ldots,L. \tag{3.5}$$

Thus, the optimum scheme allocates larger sample sizes to strata with larger variances and larger stratum sizes but smaller sample sizes to strata with larger coefficients of correlation.

3.3 Sample Size Estimation

The formula (3.5) gives n_h in terms of n, but in practice, we do not yet know what value of n is. This section presents a formula for the determination of n under the optimum sample size allocation. It is assumed that the optimum separate ratio estimate has a specified mean squared error M. When n_h, N_h and $N_h - n_h$ are all sufficient large and the technical conditions in Scott and Wu (1981) hold, we can show that the estimator $\bar{y}_{RS\_C}$ has also the asymptotic normal distribution. If the margin of error d has been given, then $M=\left(d/z_{\alpha/2}\right)^2$. Let $n_h = nw_h$, where $w_h=\frac{W_h S_h\sqrt{1-\rho_h^2}}{\sum_{h=1}^{L}W_h S_h\sqrt{1-\rho_h^2}}$. So, the mean square error of $\bar{y}_{RS\_C}$ is

$$M=\frac{1}{n}\sum_{h=1}^{L}\frac{W_h^2 S_{yh}^2\left(1-\rho_h^2\right)}{w_h}-\frac{1}{N}\sum_{h=1}^{L}W_h S_{yh}^2\left(1-\rho_h^2\right)=\left(\frac{d}{z_{\alpha/2}}\right)^2 .$$

Solving for n, we have

$$n=\frac{n_0}{1+\frac{z_{\alpha/2}^2}{Nd^2}\sum_{h=1}^{L}W_h S_h^2\sqrt{1-\rho_h^2}}, \tag{3.6}$$

where $n_0=\frac{z_{\alpha/2}^2\left(\sum_{h=1}^{L}W_h S_h\sqrt{1-\rho_h^2}\right)^2}{d^2}$.

3.4 Comparison of Efficiency

In this section, we compare the properties of the proposed optimum separate ratio estimator with the existing ones in stratified random sampling. The relative efficiency of the optimum separate ratio estimator and unbiased estimator is

$$e\left(\bar{y}_{st},\bar{y}_{RS\_C}\right)=\frac{\sum_{h=1}^{L}W_h^2\gamma_h S_{yh}^2}{\sum_{h=1}^{L}W_h^2\gamma_h S_{yh}^2\left(1-\rho_h^2\right)}. \tag{3.7}$$

This shows that the optimum separate ratio estimator is always more efficient than the unbiased estimator. The efficiency depends on the coefficient of correlation in stratum. If the correlation coefficient increases then the efficiency also increases.

To compare the properties of the optimum separate ratio estimator with the other estimators, we consider the following hypothetical populations. Each population consists of N = 50,000 units and is divided into L = 2 strata of which sizes are $N_1 = 20,000$ and $N_2 = 30,000$. The coefficients of variations are $C_{x1} = C_{y1} = 0.2$ and $C_{x2} = C_{y2} = 0.3$. The population means are given by $\overline{X}_1 = \overline{Y}_1 = 500$ and $\overline{X}_2 = \overline{Y}_2 = 1,000$. The coefficients of correlations are $\rho_h = 0.1, 0.2, \dots, 0.9$. We set the total sample size $n = 150$ with three allocations, namely equal allocation $n_h = \frac{n}{L}$, proportional allocation $n_h = \frac{nN_h}{N}$ and optimal allocation as in (3.5). The biases and MSEs of the estimators are given in Table 3 and Table 4, respectively. The biases and MSEs are computed by the formulas in the above sections.

In Table 3, the absolute biases of the unbiased and optimum separate ratio estimators are always equal to 0. The absolute biasof the estimator $\overline{y}_{RS}$ is smallest among the compared estimators. The absolute bias of the estimator $\overline{y}_{KC}$ is smaller than that of the estimator $\overline{y}_{RC}$. Given an sample size allocation, when the coefficient of correlation increases the absolute biases of the three estimators $\overline{y}_{RC}$, $\overline{y}_{RS}$, and $\overline{y}_{KC}$ decrease. Using the optimum allocation, the absolute biases of the estimators $\overline{y}_{RC}$, $\overline{y}_{RS}$ and $\overline{y}_{KC}$ are smallest among the three allocations.

Table 4 presents that the optimum separate ratio estimator gives the smallest MSE among the compared estimators. Observe that the MSE of the unbiased estimator $\overline{y}_{st}$ does not depend on the coefficient of correlation because it does not use the information of the auxillary variable. When $\rho_h > 0.5$, the MSEs of the estimators $\overline{y}_{RC}$, $\overline{y}_{RS}$ and $\overline{y}_{KC}$ are less than that of the unbiased estimator. Given an sample size allocation, when the coefficient of correlation increases, the MSEs of the estimators $\overline{y}_{RC}$, $\overline{y}_{RS}$, $\overline{y}_{KC}$ and $\overline{y}_{RS\_C}$ decrease. The MSEs of all estimators are smallest under the optimum allocation.

Table 3. Biases of the estimators in stratified random sampling

Allocation	n_1	n_2	ρ_1	ρ_2	ρ	$B(\overline{y}_{st})$	$B(\overline{y}_{RC})$	$B(\overline{y}_{RS})$	$B(\overline{y}_{KC})$	$B(\overline{y}_{RS\_C})$
equal	75	75	0.1	0.1	0.56	0	0.5087	0.4261	0.5083	0
			0.2	0.2	0.61	0	0.4522	0.3787	0.4518	0
			0.3	0.3	0.66	0	0.3957	0.3314	0.3953	0
			0.4	0.4	0.71	0	0.3391	0.2841	0.3388	0
			0.5	0.5	0.75	0	0.2826	0.2367	0.2823	0
			0.6	0.6	0.80	0	0.2261	0.1894	0.2258	0
			0.7	0.7	0.85	0	0.1696	0.1420	0.1693	0
			0.8	0.8	0.90	0	0.1130	0.0947	0.1128	0
			0.9	0.9	0.95	0	0.0565	0.0473	0.0563	0
proportion	60	90	0.1	0.1	0.56	0	0.4337	0.3709	0.4334	0
			0.2	0.2	0.61	0	0.3855	0.3297	0.3852	0
			0.3	0.3	0.66	0	0.3373	0.2885	0.3371	0
			0.4	0.4	0.71	0	0.2891	0.2473	0.2889	0
			0.5	0.5	0.75	0	0.2409	0.2060	0.2407	0
			0.6	0.6	0.80	0	0.1928	0.1648	0.1925	0
			0.7	0.7	0.85	0	0.1446	0.1236	0.1444	0
			0.8	0.8	0.90	0	0.4337	0.3709	0.4334	0
			0.9	0.9	0.95	0	0.3855	0.3297	0.3852	0
optimum	27	123	0.1	0.1	0.56	0	0.3617	0.3421	0.3614	0
			0.2	0.2	0.61	0	0.3215	0.3041	0.3213	0
			0.3	0.3	0.66	0	0.2813	0.2661	0.2811	0
			0.4	0.4	0.71	0	0.2411	0.2281	0.2409	0

	0.5	0.5	0.75	0	0.2009	0.1900	0.2007	0
	0.6	0.6	0.80	0	0.1608	0.1520	0.1606	0
	0.7	0.7	0.85	0	0.1206	0.1140	0.1204	0
	0.8	0.8	0.90	0	0.0804	0.0760	0.0802	0
	0.9	0.9	0.95	0	0.0402	0.0380	0.0400	0

ρ is the coefficient of correlation in the whole population.

Table 4. MSE of the estimators in stratified random sampling

Allocation	n_1	n_2	ρ_1	ρ_2	ρ	$MSE(\bar{y}_{st})$	$MSE(\bar{y}_{RC})$	$MSE(\bar{y}_{RS})$	$MSE(\bar{y}_{KC})$	$MSE(\bar{y}_{RS\_C})$
			0.1	0.1	0.56	452.17	813.91	813.91	813.65	447.65
			0.2	0.2	0.61	452.17	723.48	723.48	723.24	434.09
			0.3	0.3	0.66	452.17	633.04	633.04	632.84	411.48
			0.4	0.4	0.71	452.17	542.61	542.61	542.43	379.83
equal	75	75	0.5	0.5	0.75	452.17	452.17	452.17	452.03	339.13
			0.6	0.6	0.80	452.17	361.74	361.74	361.62	289.39
			0.7	0.7	0.85	452.17	271.30	271.30	271.22	230.61
			0.8	0.8	0.90	452.17	180.87	180.87	180.81	162.78
			0.9	0.9	0.95	452.17	90.43	90.43	90.41	85.91
			0.1	0.1	0.56	385.51	693.91	693.91	693.69	381.65
			0.2	0.2	0.61	385.51	616.81	616.81	616.61	370.09
			0.3	0.3	0.66	385.51	539.71	539.71	539.53	350.81
			0.4	0.4	0.71	385.51	462.61	462.61	462.46	323.83
proportion	60	90	0.5	0.5	0.75	385.51	385.51	385.51	385.38	289.13
			0.6	0.6	0.80	385.51	308.41	308.41	308.31	246.72
			0.7	0.7	0.85	385.51	231.30	231.30	231.23	196.61
			0.8	0.8	0.90	385.51	154.20	154.20	154.15	138.78
			0.9	0.9	0.95	385.51	77.10	77.10	77.08	73.25
			0.1	0.1	0.56	321.51	578.71	578.71	578.52	318.29
			0.2	0.2	0.61	321.51	514.41	514.41	514.24	308.65
			0.3	0.3	0.66	321.51	450.11	450.11	449.96	292.57
			0.4	0.4	0.71	321.51	385.81	385.81	385.68	270.07
optimum	27	123	0.5	0.5	0.75	321.51	321.51	321.51	321.40	241.13
			0.6	0.6	0.80	321.51	257.21	257.21	257.12	205.76
			0.7	0.7	0.85	321.51	192.90	192.90	192.84	163.97
			0.8	0.8	0.90	321.51	128.60	128.60	128.56	115.74
			0.9	0.9	0.95	321.51	64.30	64.30	64.28	61.09

ρ is the coefficient of correlation in the whole population.

4. Discussion

In simple random sampling, the optimum ratio estimator and its variance estimate depend on the coefficient of variation, the coefficient of correlation and the mean of the auxiliary variable in the whole population. Similarly, the optimum separate ratio estimator and its variance estimate are in terms of the coefficients of variation, the coefficient of correlation and the means of the auxiliary variable in all strata. In practice, sample estimates of these parameters may be used to substitute in the formulas of these estimates.

In simple random sampling, the relative efficiency of the optimum ratio estimator and the unbiased estimator depends on the coefficient of correlation ρ. When the coefficient of correlation between the study and auxillary variable is weak, then the relative efficiency will be low so that the unbiased estimator is almost as good as the ratio estimators. For example, if $\rho \le 0.2$, then $e(\bar{y}, \bar{y}_{c^*}) \le 1.042$. This means that if we increase the sample size about 4.2%, the unbiased estimator will have the most efficiency among the estimators in the class of $\left\{ \bar{y}_c = \bar{y} \frac{\bar{x} + c}{\bar{X} + c} : -\infty < c < \infty \right\}$. Therefore, in case of $\rho \le 0.2$ we suggest using the unbiased estimator because it uses only the information of the study variable and we do not need to collect the data of the auxillary variables. In stratified random sampling, when $\mathrm{Max}\{\rho_h\} < 0.2$, we also recommend the unbiased estimator.

For future studies, we can consider applying the ratio estimator in adaptive sampling schemes as suggested by Thompson (1990) and Sangngam (2013) for.

Acknowledgements

This work was supported by the Faculty of Science, Silpakorn University, Nakorn Pathom, Thailand.

References

Cochran, W. G. (1977). *Sampling Techniques.* Newyork: Jonh Wiley and Sons.

Hajek, J. (1960). Limiting distributions in simple random sampling from a finite population. *Publication of the Mathematical Institute of the Hungarian Academy of Scienees, 5*, 361-374.

Kadilar, C., & Cingi, H. (2003). Ratio estimators in stratified random sampling. *Biometrical Journal, 45*(2), 218–225. http://dx.doi.org/10.1002/bimj.200390007

Kadilar, C., & Cingi, H. (2005). A new ratio estimator in stratified random sampling. *Communication in Statistics Theory and Method, 34*, 597-620. http://dx.doi.org/10.1081/STA-200052156

Murthy, M. N. (1964). Product method of estimation. *Sankhya, 26*, 69-74.

Prasad, B. (1989). Some improved ratio type estimators of population mean and ratio infinite population sample surveys. *Communication in Statistics Theory and Method, 18*(1), 379–392. http://dx.doi.org/10.1080/03610928908829905

Sampath. S (2005). *Sampling Theory and Methods*. Alpha Science International Ltd. Harrow, U.K.

Sangngam, P. (2013). Unequal Probability Inverse Adaptive Cluster Sampling. *Chiang Mai Journal of Science. 40*(4), 736-742.

Scott, A., & Wu, C. (1981). On the asymptotic distribution of ratio and regression estimators.*Journal of American Statistical Associtation, 76*(373), 98-102. http://dx.doi.org/10.1080/01621459.1981.10477612

Singh, H. P., & Tailor, R. (2003). Use of known correlation coefficient in estimating the finite population mean. *Statistics in Transition, 6*, 555–560.

Sisodia, B. V. S., & Dwivedi, V. K. (1981). A modified ratio estimator using coefficient of variation of auxiliary variable. *Journal of Indian Society of Agricultural Statistics, 33*, 13–18.

Thompson, S. K. (1990). Adaptive cluster sampling. *Journal of the American Statistical Association, 85*, 1050-1059.

Upadhyaya, L. N., & Singh, H. P. (1999). Use of transformed auxiliary variable in estimating the finite population mean. *Biometrical Journal, 41*(5), 627–636. http://dx.doi.org/10.1002/(SICI)1521-4036(199909)41:5<627::AID-BIMJ627>3.0.CO;2-W