Journal of Modern Applied Statistical Methods

Volume 10 | Issue 2 Article 4

11-1-2011

Modified Ratio and Product Estimators for Population Mean in Systematic Sampling

Housila P. Singh
Vikram University, Ujjain, India, hpsujn@rediffmail.com

Rajesh Tailor
Vikram University, tailorraj@gmail.com

Narendra Kumar Jatwa
Vikram University, jatwanarendra@gmail.com

Recommended Citation
Singh, Housila P.; Tailor, Rajesh; and Jatwa, Narendra Kumar (2011) "Modified Ratio and Product Estimators for Population Mean in Systematic Sampling," *Journal of Modern Applied Statistical Methods*: Vol. 10 : Iss. 2 , Article 4.
DOI: 10.22237/jmasm/1320120180

Journal of Modern Applied Statistical Methods
November 2011, Vol. 10, No. 2, 424-435

1538 – 9472/11/$95.00

Modified Ratio and Product Estimators for Population Mean in Systematic Sampling

Housila P. Singh Rajesh Tailor Narendra Kumar Jatwa
Vikram University,
Ujjain, M. P., India

The estimation of population mean in systematic sampling is explored. Properties of a ratio and product estimator that have been suggested in systematic sampling are investigated, along with the properties of double sampling. Following Swain (1964), the cost aspect is also discussed.

Key words: Population mean, exponential estimator, systematic sampling, bias, mean squared error.

Introduction

Systematic sampling is one of the simplest sampling procedures adopted in practice and is operationally more convenient than simple random sampling. Apart from the simplicity of its concept and execution, systematic sampling is likely to be more precise than simple random sampling and even more precise than stratified sampling under certain specific conditions. In sample surveys it is common to use of auxiliary information to increase the precision of estimates of population parameters. The ratio method of estimation is a good example in this context; the ratio method of estimation is consistent, biased and gives more reliable estimates than those based on simple averages (Cochran, 1963).

If an auxiliary variate x positively (high) correlated with the study variate y is obtained for each unit in the sample and the population mean $\bar{X}$ of the auxiliary variate x is known, the classical ratio estimator for the population mean $\bar{Y}$ of the study variate y is defined by

$$\bar{y}_R = \bar{y}\frac{\bar{X}}{\bar{x}} \tag{1.1}$$

where $\bar{y}$ and $\bar{x}$ are the sample means of the study variate y and the auxiliary variate x respectively, that is, the simple averages of y and x based on the sample.

If the auxiliary variate x is negatively (high) correlated with the study variate then the classical product estimator for population mean $\bar{Y}$ of the study variate y is defined by

$$\bar{y}_P = \bar{y}\frac{\bar{x}}{\bar{X}}, \tag{1.2}$$

which was first developed by Robson (1957) and later rediscovered by Murthy (1964).

Bahl and Tuteja (1991) suggested modified ratio and product estimators for estimating the population mean $\bar{Y}$ respectively as

$$\bar{y}_{\mathrm{Re}} = \bar{y}\exp\left(\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right) \tag{1.3}$$

and

$$\bar{y}_{Pe} = \bar{y}\exp\left(\frac{\bar{x}-\bar{X}}{\bar{x}+\bar{X}}\right). \tag{1.4}$$

Housila P. Singh is a Professor in the School of Studies in Statistics. E-mail him at: hpsujn@rediffmail.com. Rajesh Tailor is a Reader in the School of Studies in Statistics. His research interests are in the field of sampling techniques. Email him at: tailorraj@gmail.com. Narenda Kumar Jatwa is a research scholar in the School of Studies in Statistics. Email him at: jatwanarendra@gmail.com.

Under simple random sampling without replacement (SRSWOR), the variances of $\bar{y}_R$, $\bar{y}_P$, $\bar{y}_{Re}$ and $\bar{y}_{Pe}$ to the first degree of approximation are given, respectively, by

$$\mathrm{Var}\left(\bar{y}_R\right)_{random} = \left(\frac{1}{n}-\frac{1}{N}\right)\left[S_y^2 + R^2S_x^2 - 2R\rho_{xy}S_xS_y\right] \tag{1.5}$$

$$\mathrm{Var}\left(\bar{y}_P\right)_{random} = \left(\frac{1}{n}-\frac{1}{N}\right)\left[S_y^2 + R^2S_x^2 + 2R\rho_{xy}S_xS_y\right] \tag{1.6}$$

$$\mathrm{Var}\left(\bar{y}_{Re}\right)_{random} = \left(\frac{1}{n}-\frac{1}{N}\right)\left[S_y^2 + (1/4)R^2S_x^2 - \rho_{xy}S_xS_y\right] \tag{1.7}$$

and

$$\mathrm{Var}\left(\bar{y}_{Pe}\right)_{random} = \left(\frac{1}{n}-\frac{1}{N}\right)\left[S_y^2 + (1/4)R^2S_x^2 + \rho_{xy}S_xS_y\right] \tag{1.8}$$

where

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

and

$$S_x^2 = \frac{1}{(N-1)}\sum_{i=1}^{N}\left(x_i - \bar{X}\right)^2$$

are population mean squares of the study variate y and the auxiliary variate x respectively, ρ_{xy} is the correlation coefficient between x and y and $R = \frac{\bar{Y}}{\bar{X}}$.

Under the SRSWOR sampling scheme

$$Var(\bar{y}) = \left(\frac{1}{n}-\frac{1}{N}\right)S_y^2. \tag{1.9}$$

Hasel (1942) and Griffth (1945-46) found systematic sampling to be efficient and convenient in sampling from certain natural populations like forest areas for estimating the volume of timber. In the case of estimating the volume of timber the leaf area or the girth of the tree may be taken as the auxiliary variable (Swain, 1964).

The properties of the ratio estimator $\bar{y}_R$ under systematic sampling have been discussed by Swain (1964) and Shukla (1971) presented the properties of product estimator $\bar{y}_P$. This article discusses the properties of the modified ratio and product estimators $\bar{y}_{Re}$ and $\bar{y}_{Pe}$ in systematic sampling in the cases of single and double sampling and comparisons are made.

Modified Estimators in Systematic Sampling: Single Sampling

Suppose N units in the population are numbered from 1 to N in some order. To select a sample of n units, if a unit at random is taken from the first k units and every k^{th} subsequent unit, then $N = nk$. This sampling method is similar to that of selecting a cluster at random out of k clusters (each cluster containing n units), made such that i^{th} cluster contains serially numbered units $i, i+k, i+2k, \ldots, i+(n-1)k$. After sampling of n units, observe both the study variate y and auxiliary variate x. Let y_{ij} *and* x_{ij} denote the observations regarding the variate y and variate x respectively on the unit bearing the serial number $i+(j-1)k$ in the population $(i = 1,2,\ldots,k; j = 1,2,\ldots,n)$. If the i^{th} sampling unit is taken at random from the first k units, then $\bar{y}_{sy}$ and $\bar{x}_{sy}$ are defined as:

$$\bar{y}_{sy} = \bar{y}_{i.} = \frac{1}{n}\sum_{j=1}^{n} y_{ij},$$

and

$$\bar{x}_{sy} = \bar{x}_{i.} = \frac{1}{n}\sum_{j=1}^{n} x_{ij}.$$

Suggested Estimator

Assuming the population mean $\overline{X}$ of the auxiliary variate x is known, Swain (1964) suggested the ratio estimator of population mean $\overline{Y}$ of the study variate y based on the systematic samples as

$$\overline{y}_{Rsy} = \overline{y}_{sy}\frac{\overline{X}}{\overline{x}_{sy}} \tag{2.1}$$

and Shukla (1971) proposed the product estimator based on systematic samples as

$$\overline{y}_{Psy} = \overline{y}_{sy}\frac{\overline{x}_{sy}}{\overline{X}}. \quad 2.2)$$

The variances of $\overline{y}_{sy}$ and $\overline{x}_{sy}$ are given approximately by

$$Var(\overline{y}_{sy}) = \left(\frac{N-1}{N}\right)\frac{S_y^2}{n}\{1+\rho_y(n-1)\} \tag{2.3}$$

where S_y^2 is the population mean square for the variate y and ρ_y is the intra-class correlation between the units of a cluster corresponding to the y variate and is given by

$$\rho_y = \frac{E(y_{ij}-\overline{Y})(y_{ij'}-\overline{Y})}{E(y_{ij}-\overline{Y})^2}$$
$$= \left[\left\{\frac{1}{kn(n-1)}\sum_{i=1}^{k}\sum_{j\neq j'=1}^{n}(y_{ij}-\overline{Y})(y_{ij'}-\overline{Y})\right\}\left\{\frac{kn}{(kn-1)S_y^2}\right\}\right] \tag{2.4}$$

and

$$Var(\overline{x}_{sy}) = \left(\frac{N-1}{N}\right)\frac{S_x^2}{n}\{1+\rho_x(n-1)\} \tag{2.5}$$

where S_x^2 and ρ_x bear the same meanings as for the study variate $y's$.

For large N, the variances of $\overline{y}_{Rsy}$ and $\overline{y}_{Psy}$ to the first degree of approximation are respectively given by

$$\begin{aligned}Var(\overline{y}_{Rsy}) &= \frac{1}{n}\left[S_y^2 + R^2S_x^2 - 2R\rho_{xy}S_xS_y\right]\\ &+\frac{1}{n}\Big[\rho_y(n-1)S_y^2 + R^2\rho_x(n-1)S_x^2\\ &-2R\rho_{xy}S_xS_y\left\{\sqrt{\{1+\rho_y(n-1)\}\{1+\rho_x(n-1)\}}-1\right\}\Big],\end{aligned} \tag{2.6}$$

and

$$\begin{aligned}Var(\overline{y}_{Psy}) &= \frac{1}{n}\left[S_y^2 + R^2S_x^2 + 2R\rho_{xy}S_xS_y\right]\\ &+\frac{1}{n}\Big[\rho_y(n-1)S_y^2 + R^2\rho_x(n-1)S_x^2\\ &+2R\rho_{xy}S_xS_y\left\{\sqrt{\{1+\rho_y(n-1)\}\{1+\rho_x(n-1)\}}-1\right\}\Big],\end{aligned} \tag{2.7}$$

Assuming the intraclass correlation to be the same for both the variates y and x, for example, $\rho_y = \rho_x = \rho$, then the $Var(\overline{y}_{Rsy})$ and $Var(\overline{y}_{Psy})$ respectively reduce to

$$\begin{aligned}Var(\overline{y}_{Rsy}) &= \frac{1}{n}\left[S_y^2 + R^2S_x^2 - 2R\rho_{xy}S_xS_y\right]\\ &+\frac{\rho(n-1)}{n}\left[S_y^2 + R^2S_x^2 - 2R\rho_{xy}S_xS_y\right]\\ &= Var(\overline{y}_R)_{random}\{1+\rho(n-1)\}\end{aligned} \tag{2.8}$$

and

$$\begin{aligned}Var(\overline{y}_{Psy}) &= \frac{1}{n}\left[S_y^2 + R^2S_x^2 + 2R\rho_{xy}S_xS_y\right]\\ &+\frac{\rho(n-1)}{n}\left[S_y^2 + R^2S_x^2 + 2R\rho_{xy}S_xS_y\right]\\ &= Var(\overline{y}_P)_{random}\{1+\rho(n-1)\}\end{aligned} \tag{2.9}$$

Following Bahl and Tuteja (1991), the following modified ratio and product estimators for population mean $\overline{Y}$ are defined respectively as

$$\bar{y}_{Re\,sy} = \bar{y}_{sy} \exp\left(\frac{\bar{X} - \bar{x}_{sy}}{\bar{X} + \bar{x}_{sy}}\right) \quad (2.10)$$

and

$$\bar{y}_{Pesy} = \bar{y}_{sy} \exp\left(\frac{\bar{x}_{sy} - \bar{X}}{\bar{x}_{sy} + \bar{X}}\right). \quad (2.11)$$

To obtain the biases and variances of the estimators $\bar{y}_{Re\,sy}$ and $\bar{y}_{Pesy}$, $\bar{y}_{sy} = \bar{Y}(1+e_0)$, $\bar{x}_{sy} = \bar{X}(1+e_1)$, is written such that $E(e_0) = E(e_1) = 0$ and

$$\left.\begin{aligned}
E(e_0^2) &= \frac{Var(\bar{y}_{sy})}{\bar{Y}^2} = \left(\frac{N-1}{N}\right)\left(\frac{C_y^2}{n}\right)\{1+(n-1)\rho_y\},\\
E(e_1^2) &= \frac{Var(\bar{x}_{sy})}{\bar{X}^2} = \left(\frac{N-1}{N}\right)\left(\frac{C_x^2}{n}\right)\{1+(n-1)\rho_x\},\\
E(e_0 e_1) &= \frac{Cov(\bar{x}_{sy}, \bar{y}_{sy})}{\bar{X}\bar{Y}} = \left(\frac{N-1}{N}\right)\left(\frac{\rho_{xy} C_x C_y}{n}\right)\sqrt{\{1+(n-1)\rho_y\}\{1+(n-1)\rho_x\}}
\end{aligned}\right\} \quad (2.12)$$

where $C_y = S_y / \bar{Y}$ and $C_x = S_x / \bar{X}$ are the population coefficients of variation of y and x respectively.

$$\begin{aligned}
\bar{y}_{Re\,sy} &= \bar{Y}(1+e_0)\exp\left\{-\frac{e_1}{(2+e_1)}\right\}\\
&= \bar{Y}(1+e_0)\exp\left\{-\frac{e_1}{2}\left(1+\frac{e_1}{2}\right)^{-1}\right\}\\
&= \bar{Y}(1+e_0)\left[1-\frac{e_1}{2}\left(1+\frac{e_1}{2}\right)^{-1}+\frac{e_1^2}{8}\left(1+\frac{e_1}{2}\right)^{-2}-\ldots\right]\\
&= \bar{Y}(1+e_0)\left[\begin{aligned}&1-\frac{e_1}{2}\left(1-\frac{e_1}{2}+\frac{e_1^2}{8}-\ldots\right)\\ &+\frac{e_1^2}{8}\left(1-e_1+\frac{3}{8}e_1^2-\ldots\right)-\ldots\end{aligned}\right]\\
&= \bar{Y}(1+e_0)\left[1-\frac{e_1}{2}+\frac{3}{8}e_1^2-\ldots\right]\\
&= \bar{Y}\left[1+e_0-\frac{e_1}{2}-\frac{e_0 e_1}{2}+\frac{3}{8}e_1^2-\ldots\right]
\end{aligned}$$

or

$$\left(\bar{y}_{Re\,sy} - \bar{Y}\right) \cong \bar{Y}\left[e_0 - \frac{e_1}{2} + \frac{3}{8}e_1^2 - \frac{e_0 e_1}{2}\right]. \quad (2.13)$$

Taking the expectations of both sides in (2.13) and using the results given by (2.12) the bias of the ratio estimator $\bar{y}_{Re\,sy}$ to the first degree of approximation is obtained as

$$\begin{aligned}
B(\bar{y}_{Re\,sy}) &= \left(\frac{N-1}{nN}\right)\bar{Y}\left[\begin{aligned}&\frac{3}{8}C_x^2\{1+\rho_x(n-1)\}\\ &-\frac{1}{2}\rho_{xy} C_x C_y\sqrt{\{1+\rho_y(n-1)\}\{1+\rho_x(n-1)\}}\end{aligned}\right]\\
&= \left(\frac{N-1}{nN}\right)\bar{Y}\left(\frac{C_x^2}{8}\right)\left[\begin{aligned}&3\{1+\rho_x(n-1)\}\\ &-4c\sqrt{\{1+\rho_y(n-1)\}\{1+\rho_x(n-1)\}}\end{aligned}\right]\\
&= \left(\frac{N-1}{nN}\right)\bar{Y}\left(\frac{C_x^2}{8}\right)\{1+\rho_x(n-1)\}\left[3-4c\sqrt{\frac{\{1+\rho_y(n-1)\}}{\{1+\rho_x(n-1)\}}}\right]
\end{aligned} \quad (2.14)$$

where $c = \rho_{xy} C_y / C_x$.

Squaring both sides of (2.13) and neglecting terms of $e's$ having power greater than two results in

$$\left(\bar{y}_{\text{Re}\,sy} - \bar{Y}\right)^2 = \bar{Y}^2 \left(e_0^2 + \frac{e_1^2}{4} - e_0 e_1 \right) \tag{2.15}$$

Taking the expectations of both sides in (2.15) and using the results given by (2.12) provides the variance of the modified ratio estimator $\bar{y}_{\text{Re}\,sy}$ as

$$\mathrm{Var}\left(\bar{y}_{\text{Re}\,sy}\right) = \left(\frac{N-1}{nN}\right)\bar{Y}^2 \begin{bmatrix} \left\{1+\rho_y(n-1)\right\}C_y^2 \\ +\left\{1+\rho_x(n-1)\right\}\dfrac{C_x^2}{4} \\ -\rho_{xy}C_xC_y\sqrt{\left\{1+\rho_y(n-1)\right\}\left\{1+\rho_x(n-1)\right\}} \end{bmatrix}$$

For large N, the above expression reduces to

$$\mathrm{Var}\left(\bar{y}_{\text{Re}\,sy}\right) = \frac{S_y^2}{n} \begin{bmatrix} \left\{1+\rho_y(n-1)\right\}S_y^2 \\ +\left\{1+\rho_x(n-1)\right\}R^2\dfrac{S_x^2}{4} \\ -R\rho_{xy}S_xS_y\sqrt{\left\{1+\rho_y(n-1)\right\}\left\{1+\rho_x(n-1)\right\}} \end{bmatrix} \tag{2.16}$$

and in the case where $\rho_y = \rho_x = \rho$, it reduces to

$$\mathrm{Var}\left(\bar{y}_{\text{Re}\,sy}\right) = \frac{1}{n}\left[S_y^2 + \frac{1}{4}R^2S_x^2 - R\rho_{xy}S_xS_y\right]\left\{1+\rho(n-1)\right\}. \tag{2.17}$$

From (1.7) and (2.17):

$$Var\left(\bar{y}_{\text{Re}\,sy}\right) = Var\left(\bar{y}_{\text{Re}}\right)_{random}\left\{1+\rho(n-1)\right\}. \tag{2.18}$$

The efficiency of the modified ratio method of estimation using systematic samples with respect to modified ratio method of estimation using sample random sampling is

$$\frac{Var\left(\bar{y}_{\text{Re}}\right)_{random}}{Var\left(\bar{y}_{\text{Re}\,sy}\right)} = \frac{1}{\left\{1+\rho(n-1)\right\}}. \tag{2.19}$$

As expected, the ratio method of estimation with systematic samples will be more efficient if $\rho < 0$. The minimum value that ρ can take is $-\left(\dfrac{1}{n-1}\right)$, when the reduction in variance is 100%.

Further expressing (2.11) in terms of $e's$:

$$\begin{aligned} \bar{y}_{Pesy} &= \bar{Y}(1+e_0)\exp\left\{\frac{e_1}{2}\left(1+\frac{e_1}{2}\right)^{-1}\right\} \\ &= \bar{Y}(1+e_0)\left[1+\frac{e_1}{2}\left(1+\frac{e_1}{2}\right)^{-1}+\frac{e_1^2}{8}\left(1+\frac{e_1}{2}\right)^{-2}+\ldots\right] \\ &= \bar{Y}(1+e_0)\left[1+\frac{e_1}{2}\left(1-\frac{e_1}{2}+\ldots\right)+\frac{e_1^2}{8}(1-e_1+\ldots)-\ldots\right] \\ &= \bar{Y}(1+e_0)\left[1+\frac{e_1}{2}-\frac{1}{8}e_1^2-\ldots\right] \\ &= \bar{Y}\left[1+e_0+\frac{e_1}{2}+\frac{e_0e_1}{2}-\frac{1}{8}e_1^2+\ldots\right] \end{aligned}$$

or

$$\left(\bar{y}_{Pesy} - \bar{Y}\right) \cong \bar{Y}\left[e_0 + \frac{e_1}{2} + \frac{e_0e_1}{2} - \frac{1}{8}e_1^2\right]. \tag{2.20}$$

Taking the expectations of both sides of (2.20) and using the results given by (2.12) provides the bias of the product estimator $\bar{y}_{Pesy}$ to the first degree of approximation as

$$B(\bar{y}_{Pesy}) =$$

$$= \left(\frac{N-1}{nN}\right)\frac{\bar{Y}}{8}\left[\begin{array}{l} 4\rho_{xy}C_xC_y\sqrt{\{1+\rho_y(n-1)\}\{1+\rho_x(n-1)\}} \\ -C_x^2\{1+\rho_x(n-1)\} \end{array}\right]$$

$$= \left(\frac{N-1}{nN}\right)\left(\frac{\bar{Y}}{8}\right)C_x^2\{1+\rho_x(n-1)\}\left[4c\sqrt{\frac{\{1+\rho_y(n-1)\}}{\{1+\rho_x(n-1)\}}}-1\right]$$

(2.21)

Squaring both sides of (2.19) and neglecting terms of $e's$ having power greater than two results in:

$$\left(\bar{y}_{Pesy}-\bar{Y}\right)^2 = \bar{Y}^2\left(e_0^2+\frac{e_1^2}{4}+e_0e_1\right). \quad (2.22)$$

Taking the expectations of both sides in (2.22) and using the results given by (2.12) provides the variance of the modified product estimator $\bar{y}_{Pesy}$ as:

$$Var(\bar{y}_{Pesy}) = \left(\frac{N-1}{nN}\right)\bar{Y}^2\left[\begin{array}{l} \{1+\rho_y(n-1)\}C_y^2 \\ +\{1+\rho_x(n-1)\}\frac{C_x^2}{4} \\ +\rho_{xy}C_xC_y\sqrt{\{1+\rho_y(n-1)\}\{1+\rho_x(n-1)\}} \end{array}\right].$$

For large N, this expression reduces to

$$Var(\bar{y}_{Pesy}) = \frac{S_y^2}{n}\left[\begin{array}{l} \{1+\rho_y(n-1)\}S_y^2 \\ +\{1+\rho_x(n-1)\}R^2\frac{S_x^2}{4} \\ +R\rho_{xy}S_xS_y\sqrt{\{1+\rho_y(n-1)\}\{1+\rho_x(n-1)\}} \end{array}\right]. \quad (2.23)$$

In the casewhere $\rho_y = \rho_x = \rho$, the expression (2.23) reduces to

$$Var(\bar{y}_{Pesy}) = \frac{1}{n}\bar{Y}^2[C_y^2+\left[\frac{1}{4}\right]C_x^2+\rho_{xy}]\{1+\rho(n-1)\}S$$

$$Var(\bar{y}_{Pesy}) = \frac{1}{n}\left[S_y^2+\frac{1}{4}R^2S_x^2+R\rho_{xy}S_xS_y\right]\{1+\rho(n-1)\} \quad (2.24)$$

From (1.8) and (2.24):

$$Var(\bar{y}_{Pesy}) = Var(\bar{y}_{Pe})_{random}\{1+\rho(n-1)\}. \quad (2.25)$$

The efficiency of the modified product method of estimation using systematic samples with respect to modified product method of estimation using sample random sampling is

$$\frac{Var(\bar{y}_{Pe})_{random}}{Var(\bar{y}_{Pesy})} = \frac{1}{\{1+\rho(n-1)\}} \quad (2.26)$$

which is greater than unity if:

$$\begin{array}{l} Var(\bar{y}_{Pe})_{random} > Var(\bar{y}_{Pesy}), \\ \frac{1}{1+\rho(n-1)} > 1, \\ \text{i.e., if } 1 > 1+\rho(n-1), \\ \text{i.e., if } 0 > \rho(n-1), \\ \text{i.e., if } \rho < 0. \end{array} \quad (2.27)$$

Thus, the modified product method of estimation using systematic samples will be more efficient than the modified product method of estimation with simple random samples if $\rho < 0$. The minimum value that ρ can take is $-\left(\frac{1}{n-1}\right)$ and, in this case, $Var(\bar{y}_{Pesy}) = 0$, that is, the reduction in variance of $\bar{y}_{Pesy}$ is 100%.

Comparison of Modified Ratio $\bar{y}_{\mathrm{Re}sy}$ and Product $\bar{y}_{Pesy}$ Estimators with Usual Unbiased Estimator $\bar{y}_{sy}$, Ratio Estimator $\bar{y}_{Rsy}$ and Product Estimator $\bar{y}_{Psy}$

For large N, the variance of $\bar{y}_{sy}$ in (2.3) reduces to

$$Var(\bar{y}_{sy}) = \frac{S_y^2}{n}\{1+\rho_y(n-1)\}. \quad (2.28)$$

From (2.16) and (2.28)

$$\begin{aligned}
&\mathrm{Var}(\bar{y}_{\mathrm{Re}sy}) - \mathrm{Var}(\bar{y}_{sy}) \\
&= \frac{1}{4n}\left[\begin{array}{l}\{1+\rho_x(n-1)\}R^2S_x^2 \\ -4R\rho_{xy}S_xS_y\sqrt{\{1+\rho_y(n-1)\}\{1+\rho_x(n-1)\}}\end{array}\right] \\
&= \frac{R^2S_x^2\{1+\rho_x(n-1)\}}{4n}\left[1-4\rho_{xy}\frac{S_y}{S_xR}\sqrt{\frac{\{1+\rho_y(n-1)\}}{\{1+\rho_x(n-1)\}}}\right] \\
&= \frac{R^2S_x^2\{1+\rho_x(n-1)\}}{4n}\left[1-4\frac{\beta}{R}\sqrt{\frac{\{1+\rho_y(n-1)\}}{\{1+\rho_x(n-1)\}}}\right] \\
&= \frac{R^2S_x^2\{1+\rho_x(n-1)\}}{4n}\left[1-4c\sqrt{\frac{\{1+\rho_y(n-1)\}}{\{1+\rho_x(n-1)\}}}\right]
\end{aligned}$$

which is negative if:

$$1-4c\sqrt{\frac{\{1+\rho_y(n-1)\}}{\{1+\rho_x(n-1)\}}}<0,$$

$$\text{i.e., if } \frac{1}{4}<c\sqrt{\frac{\{1+\rho_y(n-1)\}}{\{1+\rho_x(n-1)\}}},$$

$$\text{i.e., if } c\sqrt{\frac{\{1+\rho_y(n-1)\}}{\{1+\rho_x(n-1)\}}}>\frac{1}{4},$$

$$\text{i.e., if } c>\frac{1}{4}\sqrt{\frac{\{1+\rho_x(n-1)\}}{\{1+\rho_y(n-1)\}}}, \quad (2.29)$$

where $\beta = \rho_{xy}\frac{S_y}{S_x}$ is the population regression coefficient of y on x and $c = \frac{\beta}{R}$.

From (2.6) and (2.16)

$$\begin{aligned}
&\mathrm{Var}(\bar{y}_{\mathrm{Re}sy}) - \mathrm{Var}(\bar{y}_{Rsy}) \\
&= \frac{S_y^2}{n}\left[\begin{array}{l}-\frac{3}{4}R^2S_x^2\{1+\rho_x(n-1)\} \\ +R\rho_{xy}S_xS_y\sqrt{\{1+\rho_y(n-1)\}\{1+\rho_x(n-1)\}}\end{array}\right] \\
&= \left(\frac{R^2S_x^2S_y^2}{n}\right)\{1+\rho_x(n-1)\}\left[\frac{\beta}{R}\sqrt{\frac{\{1+\rho_y(n-1)\}}{\{1+\rho_x(n-1)\}}}-\frac{3}{4}\right] \\
&= \left(\frac{R^2S_x^2S_y^2}{n}\right)\{1+\rho_x(n-1)\}\left[c\sqrt{\frac{\{1+\rho_y(n-1)\}}{\{1+\rho_x(n-1)\}}}-\frac{3}{4}\right]
\end{aligned}$$

which is negative if:

$$c\sqrt{\frac{\{1+\rho_y(n-1)\}}{\{1+\rho_x(n-1)\}}}-\frac{3}{4}<0$$

$$\text{i.e., if } c<\frac{3}{4}\sqrt{\frac{\{1+\rho_x(n-1)\}}{\{1+\rho_y(n-1)\}}}. \quad (2.30)$$

Thus, from (2.29) and (2.30) it follows that the modified ratio estimator $\bar{y}_{\mathrm{Re}sy}$ is more efficient than usual unbiased estimator $\bar{y}_{sy}$ and Swain's (1964) estimator $\bar{y}_{Rsy}$ if:

$$\frac{1}{4}\sqrt{\frac{\{1+\rho_x(n-1)\}}{\{1+\rho_y(n-1)\}}}<c<\frac{3}{4}\sqrt{\frac{\{1+\rho_x(n-1)\}}{\{1+\rho_y(n-1)\}}}$$

i.e., if

$$\frac{1}{4}\frac{C_x}{C_y}\sqrt{\frac{\{1+\rho_x(n-1)\}}{\{1+\rho_y(n-1)\}}}<\rho_{xy}<\frac{3}{4}\frac{C_x}{C_y}\sqrt{\frac{\{1+\rho_x(n-1)\}}{\{1+\rho_y(n-1)\}}} \quad (2.31)$$

when the intraclass correlation coefficients for both the variates are same (i.e. $\rho_y = \rho_x = \rho$), then condition (2.31) reduces to:

$$\frac{1}{4}\frac{C_x}{C_y} < \rho_{xy} < \frac{3}{4}\frac{C_x}{C_y}. \qquad (2.32)$$

From (2.7), (2.23) and (2.28)

$$\mathrm{Var}\left(\bar{y}_{Pesy}\right) - \mathrm{Var}\left(\bar{y}_{sy}\right) = \left(\frac{R^2 S_x^2 S_y^2}{n}\right)\left\{1+\rho_x(n-1)\right\}\left[\frac{1}{4} + c\sqrt{\frac{\left\{1+\rho_y(n-1)\right\}}{\left\{1+\rho_x(n-1)\right\}}}\right] \qquad (2.33)$$

and

$$\mathrm{Var}\left(\bar{y}_{Pesy}\right) - \mathrm{Var}\left(\bar{y}_{Psy}\right) = \left(\frac{R^2 S_x^2 S_y^2}{n}\right)\left\{1+\rho_x(n-1)\right\}\left[-\frac{3}{4} - c\sqrt{\frac{\left\{1+\rho_y(n-1)\right\}}{\left\{1+\rho_x(n-1)\right\}}}\right] \qquad (2.34)$$

It follows from (2.33) and (2.34) respectively that the proposed modified product estimator $\bar{y}_{Pesy}$ is more efficient than

(i) usual unbiased estimator $\bar{y}_{sy}$ if

$$Var\left(\bar{y}_{Pesy}\right) < Var\left(\bar{y}_{sy}\right),$$

$$\text{i.e., if } \frac{1}{4} + c\sqrt{\frac{\left\{1+\rho_y(n-1)\right\}}{\left\{1+\rho_x(n-1)\right\}}} < 0,$$

$$\text{i.e., if } c < -\frac{1}{4}\sqrt{\frac{\left\{1+\rho_x(n-1)\right\}}{\left\{1+\rho_y(n-1)\right\}}}. \qquad (2.35)$$

(ii) Shukla's (1971) product estimator $\bar{y}_{Psy}$ if

$$Var\left(\bar{y}_{Pesy}\right) < Var\left(\bar{y}_{Psy}\right),$$

$$\text{i.e., if } -\frac{3}{4} - c\sqrt{\frac{\left\{1+\rho_y(n-1)\right\}}{\left\{1+\rho_x(n-1)\right\}}} < 0,$$

$$\text{i.e., if } -\frac{3}{4} < c\sqrt{\frac{\left\{1+\rho_y(n-1)\right\}}{\left\{1+\rho_x(n-1)\right\}}},$$

$$\text{i.e., if } c > -\frac{3}{4}\sqrt{\frac{\left\{1+\rho_x(n-1)\right\}}{\left\{1+\rho_y(n-1)\right\}}}. \qquad (2.36)$$

Thus, from (2.35) and (2.36) it follows that the proposed modified product estimator $\bar{y}_{Pesy}$ is more efficient than usual unbiased estimator $\bar{y}_{sy}$ and Shukla's (1971) estimator $\bar{y}_{Psy}$ if

$$-\frac{3}{4}\sqrt{\frac{\left\{1+\rho_x(n-1)\right\}}{\left\{1+\rho_y(n-1)\right\}}} < c < -\frac{1}{4}\sqrt{\frac{\left\{1+\rho_x(n-1)\right\}}{\left\{1+\rho_y(n-1)\right\}}} \qquad (2.37)$$

i.e., if

$$-\frac{3}{4}\frac{C_x}{C_y}\sqrt{\frac{\left\{1+\rho_x(n-1)\right\}}{\left\{1+\rho_y(n-1)\right\}}} < \rho_{xy} < -\frac{1}{4}\frac{C_x}{C_y}\sqrt{\frac{\left\{1+\rho_x(n-1)\right\}}{\left\{1+\rho_y(n-1)\right\}}} \qquad (2.38)$$

In the case where $\rho_y = \rho_x = \rho$, the condition (2.38) reduces to:

$$-\frac{3}{4}\frac{C_x}{C_y} < \rho_{xy} < -\frac{1}{4}\frac{C_x}{C_y}. \qquad (2.39)$$

Modified Estimators in Systematic Sampling: Two-Phase (or Double) Sampling

If the population mean $\bar{X}$ of the auxiliary variable x is not known before start of the survey, then it may be more efficient to conduct the sampling in two-phase (or double) sampling by taking a large preliminary sample to estimate the population mean $\bar{X}$. This method is a powerful and cost effective (economical) procedure and, therefore, has role to play in survey sampling (Hidiroglou & Sarndal, 1998; Hidiroglou, 2001).

In the present situation the population is divided into k clusters of n units each according to the previous rule and λ clusters (λ being less than k) and selected to observe only the auxiliary variate x, while another cluster is selected to observe both y and x variates (Swain, 1964). If $\bar{x}$ is the mean of the $x's$ from the selected λ clusters, then

$$\bar{x}' = \frac{1}{\lambda n}\sum_{i=1}^{\lambda}\sum_{j=1}^{n} x_{ij}, \qquad (3.1)$$

such that $E(\bar{x}') = \bar{X}$, that is, $\bar{x}'$ is an unbiased estimator of the population mean $\bar{X}$.Swain (1964) suggested the double sampling ratio estimator with systematic samples as

$$\bar{y}_{Rsy}^{(d)} = \bar{y}_{sy}\left(\frac{\bar{x}'}{\bar{x}_{sy}}\right). \qquad (3.2)$$

The double sampling version of product estimator $\bar{y}_{Psy}$ in (2.29) is defined by

$$\bar{y}_{Psy}^{(d)} = \bar{y}_{sy}\left(\frac{\bar{x}_{sy}}{\bar{x}'}\right). \qquad (3.3)$$

Case I

For large N, $\rho_y = \rho_x = \rho$ and, if the first set of λ clusters and the second cluster are chosen randomly and independently, the variances of the double sampling ratio $\left(\bar{y}_{Rsy}^{(d)}\right)$ and product $\left(\bar{y}_{Psy}^{(d)}\right)$ based on systematic samples to the first degree of approximation are respectively given by

$$\mathrm{Var}\left(\bar{y}_{Rsy}^{(d)}\right) = \frac{1}{n}\begin{bmatrix} S_y^2 + R^2 S_x^2 \\ -2R\rho_{xy}S_yS_x \end{bmatrix}\{1+\rho(n-1)\} + \frac{R^2S_x^2}{\lambda n}\{1+\rho(n-1)\} \qquad (3.4)$$

and

$$\mathrm{Var}\left(\bar{y}_{Psy}^{(d)}\right) = \frac{1}{n}\begin{bmatrix} S_y^2 + R^2 S_x^2 \\ +2R\rho_{xy}S_yS_x \end{bmatrix}\{1+\rho(n-1)\} + \frac{R^2S_x^2}{\lambda n}\{1+\rho(n-1)\}. \qquad (3.5)$$

Case II

For large N, $\rho_y = \rho_x = \rho$ and if the second cluster is chosen randomly from the first set of selected clusters, the variances of the double sampling ratio and product estimators based on systematic sampling are respectively given by

$$\mathrm{Var}\left(\bar{y}_{Rsy}^{(d)}\right) = \frac{1}{n}\begin{bmatrix} S_y^2 + R^2 S_x^2 \\ -2R\rho_{xy}S_yS_x \end{bmatrix}\{1+\rho(n-1)\} + \frac{1}{\lambda n}\left[\begin{Bmatrix} 2R\rho_{xy}S_yS_x \\ -R^2S_x^2 \end{Bmatrix}\begin{Bmatrix} 1+ \\ \rho(n+1) \end{Bmatrix}\right] \qquad (3.6)$$

and

$$\mathrm{Var}\left(\bar{y}_{Psy}^{(d)}\right) = \frac{1}{n}\begin{bmatrix} S_y^2 + R^2 S_x^2 \\ +2R\rho_{xy}S_yS_x \end{bmatrix}\{1+\rho(n-1)\} + \frac{1}{\lambda n}\left[\begin{Bmatrix} -2R\rho_{xy}S_yS_x \\ -R^2S_x^2 \end{Bmatrix}\begin{Bmatrix} 1 \\ +\rho(n-1) \end{Bmatrix}\right] \qquad (3.7)$$

Following Bahl and Tuteja (1991) and Singh and Vishwakarma (2007) a modified double sampling ratio estimator based on systematic sampling is proposed a

$$\bar{y}_{\mathrm{Re}sy}^{(d)} = \bar{y}_{sy}\exp\left(\frac{\bar{x}'_{sy} - \bar{x}_{sy}}{\bar{x}'_{sy} + \bar{x}_{sy}}\right), \qquad (3.8)$$

and the modified double sampling product estimator based on systematic sampling

$$\bar{y}_{Pesy}^{(d)} = \bar{y}_{sy} \exp\left(\frac{\bar{x}_{sy} - \bar{x}'_{sy}}{\bar{x}_{sy} + \bar{x}'_{sy}}\right). \quad (3.9)$$

Case I

For large N, $\rho_y = \rho_x = \rho$ and if the first set of λ clusters and the second cluster is chosen randomly and independently, the variances of the modified double sampling ratio and product estimators based on systematic samples to the first degree of approximation are respectively given by

$$\mathrm{Var}\left(\bar{y}_{\mathrm{Re}sy}^{(d)}\right) = \frac{1}{n}\begin{bmatrix} S_y^2 + (1/4)R^2S_x^2 \\ -R\rho_{xy}S_yS_x \end{bmatrix}\{1+\rho(n-1)\} + \frac{R^2S_x^2}{4\lambda n}\{1+\rho(n-1)\} \quad (3.10)$$

and

$$\mathrm{Var}\left(\bar{y}_{Pesy}^{(d)}\right) = \frac{1}{n}\begin{bmatrix} S_y^2 + (1/4)R^2S_x^2 \\ +R\rho_{xy}S_yS_x \end{bmatrix}\{1+\rho(n-1)\} + \frac{R^2S_x^2}{4\lambda n}\{1+\rho(n-1)\} \quad (3.11)$$

Case II

If the second cluster is selected randomly from the first set of selected clusters, then the variances of the double sampling ratio $\bar{y}_{\mathrm{Re}sy}^{(d)}$ and product $\bar{y}_{Pesy}^{(d)}$ estimators to the first degree of approximation are respectively given by

$$\mathrm{Var}\left(\bar{y}_{\mathrm{Re}sy}^{(d)}\right) = \frac{1}{n}\left[S_y^2 + (1/4)R^2S_x^2 - R\rho_{xy}S_yS_x\right]\{1+\rho(n-1)\} + \frac{1}{\lambda n}\left[\left\{R\rho_{xy}S_yS_x - (1/4)R^2S_x^2\right\}\{1+\rho(n+1)\}\right] \quad (3.12)$$

and

$$\mathrm{Var}\left(\bar{y}_{Pesy}^{(d)}\right) = \frac{1}{n}\left[S_y^2 + (1/4)R^2S_x^2 - R\rho_{xy}S_yS_x\right]\{1+\rho(n-1)\} + \frac{1}{\lambda n}\left[\left\{-R\rho_{xy}S_yS_x - (1/4)R^2S_x^2\right\}\{1+\rho(n-1)\}\right] \quad (3.13)$$

For large N and $\rho_y = \rho$, the variance of usual unbiased estimator $\bar{y}_{sy}$ is given by

$$Var(\bar{y}_{sy}) = \frac{S_y^2}{n}\{1+\rho(n-1)\}. \quad (3.14)$$

Efficiency Comparisons

From (3.4), (3.5), (3.10), (3.11) and (3.14) in Case I it can be shown that the proposed estimator $\bar{y}_{\mathrm{Re}sy}^{(d)}$ is more efficient than

(a) the usual unbiased estimator $\bar{y}_{sy}$ if

$$\rho_{xy} > \frac{C_x}{4C_y}\left(1+\frac{1}{\lambda}\right) \quad (3.15)$$

(b) Swain's (1964) estimator $\bar{y}_{Rsy}^{(d)}$ if

$$\rho_{xy} < \frac{3C_x}{4C_y}\left(1+\frac{1}{\lambda}\right) \quad (3.16)$$

and that $\bar{y}_{Pesy}^{(d)}$ is better than

(a) the usual unbiased estimator $\bar{y}_{sy}$ if

$$\rho_{xy} < -\frac{C_x}{4C_y}\left(1+\frac{1}{\lambda}\right) \quad (3.17)$$

(b) the product estimator $\bar{y}_{Psy}^{(d)}$ if

$$\rho_{xy} > -\frac{3C_x}{4C_y}\left(1+\frac{1}{\lambda}\right). \quad (3.18)$$

Combining {(3.15) and (3.16)} and {(3.17) and (3.18)} shows that the proposed estimator $\bar{y}_{\mathrm{Re}\,sy}$ is more efficient than $\bar{y}_{sy}$ and $\bar{y}_{Rsy}^{(d)}$ if

$$\frac{C_x}{4C_y}\left(1+\frac{1}{\lambda}\right)<\rho_{xy}<\frac{3C_x}{4C_y}\left(1+\frac{1}{\lambda}\right) \quad (3.19)$$

and the proposed modified product estimator $\bar{y}_{Pesy}^{(d)}$ is better than $\bar{y}_{sy}$ and $\bar{y}_{Psy}^{(d)}$ if

$$-\frac{3C_x}{4C_y}\left(1+\frac{1}{\lambda}\right)<\rho_{xy}<-\frac{C_x}{4C_y}\left(1+\frac{1}{\lambda}\right). \quad (3.20)$$

From (3.6), (3.7), (3.12), (3.13) and (3.14) in case II it can be established that the proposed estimator $\bar{y}_{\mathrm{Re}\,sy}^{(d)}$ is better than

(a) the usual unbiased estimator $\bar{y}_{sy}$ if

$$\rho_{xy}>\frac{1}{4}\frac{C_x}{C_y} \quad (3.21)$$

(b) the ratio estimator $\bar{y}_{Rsy}^{(d)}$ if

$$\rho_{xy}<\frac{3}{4}\frac{C_x}{C_y} \quad (3.22)$$

and $\bar{y}_{Pesy}^{(d)}$ is more efficient than

(a) the usual unbiased estimator $\bar{y}_{sy}$ if

$$\rho_{xy}<-\frac{1}{4}\frac{C_x}{C_y} \quad (3.23)$$

(b) the product estimator $\bar{y}_{Psy}^{(d)}$ if

$$\rho_{xy}>-\frac{3}{4}\frac{C_x}{C_y}. \quad (3.24)$$

Combining (3.21) and (3.22) shows that the proposed estimator $\bar{y}_{\mathrm{Re}\,sy}^{(d)}$ is more efficient than $\bar{y}_{sy}$ and $\bar{y}_{Rsy}^{(d)}$ if

$$\frac{1}{4}\frac{C_x}{C_y}<\rho_{xy}<\frac{3}{4}\frac{C_x}{C_y},$$

a condition which is usually met in practice. Further from (3.23) and (3.24) it follows that the proposed estimator $\bar{y}_{Pesy}^{(d)}$ is better than $\bar{y}_{sy}$ and $\bar{y}_{Psy}^{(d)}$ if:

$$-\frac{3}{4}\frac{C_x}{C_y}<\rho_{xy}<-\frac{1}{4}\frac{C_x}{C_y}.$$

Cost Aspect

Following Swain (1964), let the cost function be of the form

$$C^{*}=c_0 n+c_1\lambda n=(c_0+c_1\lambda)n \quad (3.25)$$

where:

C^{*} = total cost,

c_0 = cost for observing a pair of (y,x) on a sampling unit, and

c_1 = cost for observing x on any unit of λ clusters.

From (3.10), (3.11), (3.12) and (3.13), note that all the four variance formulae are of the form:

$$V=\frac{V_1}{n}\{1+\rho(n-1)\}+\frac{V_2}{\lambda n}\{1+\rho(n-1)\}. \quad (3.26)$$

The optimum values of n and λ can be obtained by minimizing the variance function for a given cost. The value of λ which minimizes the variance function can be obtained by the equation

$$\frac{\partial V}{\partial \lambda}=0,$$

where

$$V = V_1\left[\frac{(c_0+c_1\lambda)}{C}\left\{1+\rho\left(\frac{C}{c_0+c_1\lambda}-1\right)\right\}\right] + V_2\left[\frac{(c_0+c_1\lambda)}{C\lambda}\left\{1+\rho\left(\frac{C}{c_0+c_1\lambda}-1\right)\right\}\right]. \quad (3.27)$$

Differentiating (3.27) with respect to λ and equating to zero results in

$$\frac{\partial V}{\partial \lambda} = 0$$

$$= V_1\frac{c_1}{C}(1-\rho) - \frac{V_2}{\lambda^2}\left\{\rho + (1-\rho)\frac{c_0}{C}\right\}$$

$$\Rightarrow V_1\frac{c_1}{C}(1-\rho) = \frac{V_2}{\lambda^2}\left\{\rho + (1-\rho)\frac{c_0}{C}\right\}$$

$$\Rightarrow \lambda^2 = \frac{V_2}{V_1}\left[\frac{c_0}{c_1} + \frac{\rho}{(1-\rho)}\frac{C}{c_1}\right]$$

which gives

$$\lambda_{opt} = \sqrt{\frac{V_2}{V_1}\left[\frac{c_0}{c_1} + \frac{\rho}{(1-\rho)}\frac{C}{c_1}\right]}. \quad (3.28)$$

Substituting (3.28) in (3.25) results in

$$n_{opt} = \frac{C}{(c_0 + c_1\lambda_{opt})} = \frac{C}{c_0 + c_1\sqrt{\frac{V_2}{V_1}\left[\frac{c_0}{c_1} + \frac{\rho}{(1-\rho)}\frac{C}{c_1}\right]}}, \quad (3.29)$$

and substitution of (2.28) and (3.29) in (3.26) yields the minimum variance

$$V_{opt} = \frac{V_1}{n_{opt}}\left\{1+\rho(n_{opt}-1)\right\} + \frac{V_2}{\lambda_{opt} n_{opt}}\left\{1+\rho(n_{opt}-1)\right\}. \quad (3.30)$$

References

Cochran, W. G. (1946). Relative accuracy of systematic and stratified random samples for a certain class of populations. *Annals of Mathematical Statistics*, *17*, 164-177.

Robson, D. S. (1957).Application of multivariate polykays to the theory of unbiased ratio-type estimation. *Journal of the American Statistical Association*, *59*, 1225-1226.

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

Bahl, S., & Tuteja, R. K. (1991). Ratio and product type exponential estimators. Information and optimization. *Science*, *12*, 159-163.

Hasel, H. A. (1942). Estimation of volume of timber stands by strip sampling. *Annals of Mathematical Statistics*, *13*, 179-206.

Griffth, A. L. (1945-46). The efficiency of enumerations. *Indian Forest Leaflets*, *83-93*. Forest-Research Institute: Dehra Dun.

Swain, A. K. P. C. (1964). The use of systematic sampling ratio estimate. *Journal of the Indian Statistical Association*, *2*,160-164.

Shukla, N. D. (1971). Systematic sampling and product method of estimation.In *Proceedings of all India Seminar on Demography and Statistics*. B.H.U. Varanasi: India.

Singh, H. P., & Vishwakarma, G. K. (2007). Modified exponential ratio and product estimators for finite population mean in double sampling. *Austrian Journal of Statistics*, *36*, 217-225.