World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(Review Article)

Check for updates

An efficient ratio estimator of population means under double sampling technique using information on auxiliary attribute

Catherine Shike Agbebia [1,*], Andrew Ekpung Ntul [1], Etorti Imoke John [2] and Egomo Eric Monfung [1]

[1] *Department of Statistics, University of Cross River State (UNICROSS), Calabar, Nigeria.*
[2] *Department of Mathematics, College of Education, Akamkpa- Nigeria.*

World Journal of Advanced Research and Reviews, 2024, 22(02), 1165–1176

Publication history: Received on 04 April 2024; revised on 13 May 2024; accepted on 15 May 2024

Article DOI: https://doi.org/10.30574/wjarr.2024.22.2.1313

Abstract

This paper provides the estimation of population mean under double sampling, using supplementary information in the form of an attribute (that is, a proportion of the population possessing an attribute highly correlated with the study variable). A generalized efficient family of ratio estimator of the population mean is suggested and expressions for the bias and the mean square error of the estimator, as well as the minimum value are obtained. The proposed estimator has an improvement over some existing estimators when compared numerically using a natural population data and therefore provides a better substitute for the purpose of estimation.

Keyword*:* Double sampling; Proportion; Efficiency; Estimator and population mean

1. Introduction

In the theory of sample survey, auxiliary information or variables plays a vital role for increasing the efficiency of estimators of population parameters when the study variable y is highly correlated with the auxiliary variable x. But in several cases, instead of the existence of auxiliary variables, there exist some auxiliary attributes which are highly correlated with the study variable y. For example, the amount of milk produced and a particular breed of cow, the yield of cassava and a particular variety of cassava etc. In such situations, taking advantage of the correlation between the study variable and the auxiliary attribute, the estimators of population parameters can be constructed by using prior knowledge of the parameter of the auxiliary attribute. Where this prior information is not feasible, doubling sampling becomes necessary.

Double sampling is useful for obtaining auxiliary variables for ratio and regression estimation and also for finding information for stratified sampling. In this scheme, a large preliminary sample is selected at the first instance from which the missing auxiliary information only is obtained. Thereafter, a second sample is selected in which the variable of interest is measured in addition to the auxiliary information. A number of authors have developed estimators based on auxiliary attributes. Naik and Gupta (1996), Singh et al (2007), Abd-Elfattah et al. (2010), Solanki and Singh (2012), Bahl and Tuteja (1991), Yadav et. al. (2013), Koyuncu (2012) etc.

Let Y_i, X_i and Θ_i be the observations on the ith unit of the population for the study variable y, the auxiliary variable x and the auxiliary attribute Θ (i=1, 2, …N). Consider the following notations

N = Population size

* Corresponding author: Agbebia Catherine Shike

n = Sample size

$A = \sum_{i=1}^{N} \Theta_i$ = The total number of units in the population possession attribute Θ

$a = \sum_{i=1}^{n} \Theta_i$ = The total number of units in the sample possession attribute Θ .

$P = \frac{A}{N}$ = The proportion of units in the population possessing attribute Θ

$p = \frac{a}{n}$ = The proportion of units in the sample possessing attribute Θ .

$S_y^2 =$ The population variance of the study variable

$S_\Theta^2 =$ The population variance of the auxiliary attribute

$S_x^2 =$ The population variance of the auxiliary variable

Suppose the information about the auxiliary attribute or variable is not known, then in double sampling, the auxiliary attribute or variable are replaced by the corresponding sample values from a large preliminary sample of size n^1 drawn using simple random sampling without replacement from a population of size N in the first phase. Also, both the study variable y and the auxiliary attribute Θ or variable X are observed on the second phase sample of size n^2 drawn from the first phase sample by simple random sampling without replacement. Let

$$p^1 = \frac{\sum_{i=1}^{n^1} \Theta_i}{n^1} = \text{ First phase sample mean of auxiliary attribute}$$

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \text{Second phase sub-sample mean of study variable}$$

$$\bar{x}^1 = \frac{\sum_{i=1}^{n^1} x_i}{n^1} = \text{ First phase sample mean of auxiliary variable}$$

$$p = \frac{\sum_{i=1}^{n^2} \Theta_i}{n} = \text{Second phase sub-sample mean of auxiliary attribute}$$

$$\bar{x} = \frac{\sum_{i=1}^{n^2} x_i}{n^2} = \text{ Second phase sub-sample mean of auxiliary variable}$$

$C_x^2 = \frac{S_x^2}{\overline{X}^2}$ = Coefficient of variation of auxiliary variable

$C_y^2 = \frac{S_y^2}{\overline{Y}^2}$ = Coefficient of variation of study variable

$C_\Theta^2 = \frac{S_\Theta^2}{P^2}$ = Coefficient of variation of auxiliary attribute

Consider the following ratio and exponential ratio-type estimators of population mean developed in the past.

General simple random sampling without replacement (SRSWOR) estimator of population mean

$\hat{\overline{Y}}_1 = \overline{y}$ (Sample mean)

$$\text{MSE}\,(\hat{\overline{Y}}_1) = \lambda \overline{Y}^2 C_y^2 \quad (1)$$

(2) Double sampling ratio estimator using auxiliary variable

$$\hat{\overline{Y}}_2 = \overline{y}\frac{\overline{x}^1}{\overline{x}}$$

$$\text{MSE}\,(\hat{\overline{Y}}_2) = \overline{Y}^2\left[\lambda C_y^2 + (\lambda - \lambda^1)\left(C_x^2 - 2\rho_{xy} C_y C_x\right)\right] \quad (2)$$

(3) Naik and Gupta (1996) double sampling ratio estimator using auxiliary attribute

$$\hat{\overline{Y}}_3 = \overline{y}\frac{p^1}{p}$$

$$\text{MSE}((\hat{\overline{Y}}_3) = \overline{Y}^2\left[\lambda C_y^2 + (\lambda - \lambda^1)\left(C_p^2 - 2\rho_{yp} C_y C_p\right)\right] \quad (3)$$

(4) Bahl and Tuteja (1991) double sampling exponential ratio estimator using auxiliary variable

$$\hat{\overline{Y}}_4 = \overline{y}\exp\left(\frac{\overline{x}^1 - \overline{x}}{\overline{x}^1 + \overline{x}}\right)$$

$$\text{MSE}\,(\hat{\overline{Y}}_4) = \overline{Y}^2\left[\lambda C_y^2 + (\lambda - \lambda^1)\frac{1}{4}\left(C_x^2 - \rho_{xy} C_y C_x\right)\right] \quad (4)$$

(5) Nirmala Sawan (2010) double sampling exponential ratio estimator using auxiliary attribute

$$\hat{\overline{Y}}_5 = \overline{y}\exp\left(\frac{p^1 - p}{p^1 + p}\right)$$

$$MSE(\hat{\overline{Y}}_5) = \overline{Y}^2\left[\lambda C_y^2 + (\lambda - \lambda^1)\frac{1}{4}\left(C_p^2 - \rho_{py}C_yC_p\right)\right] \quad (5)$$

(6) Etorti et. al (2023), generalized family of ratio estimator using auxiliary variable

$$\hat{\overline{Y}}_6 = \overline{y}\left(\frac{\overline{x}^1 + a\overline{x}}{a\overline{x}^1 + \overline{x}}\right)^\gamma$$

Case I : when the second phase sample n^2 is drawn from a large preliminary first sample n^1

$$MSE_{\min}(\hat{\overline{Y}}_6) = \overline{Y}^2 C_y^2\left[\lambda - (\lambda - \lambda^1)\rho^2{}_{xy}\right] \quad (6)$$

Case II: when the second phase sample n^2 is drawn independently from the population not the first phase sample n^1 such that $E(e_{x_1}e_{x_2}) = E(e_ye_{x_1}) = 0$

$$MSE\min(\hat{\overline{Y}}_6) = \overline{Y}^2 C_y^2\left[\lambda - w\rho^2{}_{xy}\right], \; w = \frac{\lambda^2}{\lambda + \lambda^1} \quad (7)$$

$$\lambda = \frac{1}{n^2} - \frac{1}{N}, \; \lambda^1 = \frac{1}{n^1} - \frac{1}{N}, \; (\lambda - \lambda^1) = \frac{1}{n^2} - \frac{1}{n^1}$$

P_{xy} is the correlation coefficient between x and y

2. Proposed Estimator

Motivated by Naik and Gupta (1996), Nirmala Sawan (2010) and Etorti et. al. (2023), we proposed the following new ratio-type estimator of population mean under double sampling using auxiliary attribute.

$$\hat{\overline{Y}}_7 = \overline{y}\left(\frac{p^1 + ap}{ap^1 + p}\right)^\gamma \quad (8)$$

where a and γ are suitably chosen constants. It may be noted that the estimators given in (1) and (3) are special cases of the proposed estimator when γ=0 and a=0.

In order to obtain the bias and mean square error (MSE), let us denote

$$e_y = \frac{\overline{y} - \overline{Y}}{\overline{Y}}, \; e_{p^1} = \frac{p^1 - P}{P}, \; e_p = \frac{p - P}{P} \; such\ that\ p^1 = P\left(1 + e_{p^1}\right), \; p = P\left(1 + e_p\right), \; \overline{y} = \overline{Y}\left(1 + e_y\right)$$

$$E(e_y) = E(e_p) = E(e_{p^1}) = 0 \; and \; E(e_y^2) = \lambda C_y^2, \; E(e_p^2) = \lambda C_p^2, \; E(e_{p^1}^2) = \lambda^1 C_p^2,$$

$$E(e_ye_p) = \lambda\rho_{yp}C_yC_p \; E(e_ye_{p^1}) = \lambda^1\rho_{yp}C_yC_p$$

Case I :

The proposed estimator can be expressed in terms of the e's as :

$$\hat{\bar{Y}}_7 = \bar{Y}(1+e_y)\left(\frac{1+e_{p^1}+a+ae_p}{a+ae_{p^1}+1+e_p}\right)^{\gamma} = \bar{Y}(1+e_y)(1+u)^{\gamma}(1+V)^{-\gamma}$$

u = $g_1e_{p1}+g_2e_p$, v= $g_2e_{p1}+g_1e_p$ and $g_1=\frac{1}{1+a}$, $g_2=\frac{a}{1+a}$. Assuming $|v|<1$, $(1+u)^{\gamma}$ *and* $(1+v)^{-\gamma}$ can be expanded using Taylor approximation as:

$$\hat{\bar{Y}}_7 = \bar{Y}(1+e_y)(1+u)^{\gamma}(1+V)^{-\gamma} = \bar{Y}(1+e_y)(1+\gamma u+\gamma(\frac{\gamma-1}{2})u^2+\ldots)(1-\gamma v+\gamma(\frac{\gamma+1}{2})v^2+\ldots) \quad (9)$$

Expanding (9) and retaining terms up to the second order, we get

$$\hat{\bar{Y}}_7 = \bar{Y}\left\{1+\gamma(u-v)+(\frac{\gamma^2+\gamma}{2})v^2+(\frac{\gamma^2-\gamma}{2})u^2-\gamma^2uv+e_y-\gamma e_y(v-u)\right\}$$

$$\left(\hat{\bar{Y}}_7-\bar{Y}\right) = \bar{Y}\left\{\begin{array}{l} e_y+(g_1-g_2)\gamma e_{p1}-(g_1-g_2)\gamma e_p+\left[g_1^2(\frac{\gamma^2-\gamma}{2})+g_2^2(\frac{\gamma^2+\gamma}{2})-\gamma^2g_1g_2\right]e_{p1}^2+ \\ \left[-\gamma 2g_1g_2+g_1^2(\frac{\gamma^2-\gamma}{2})+g_2^2(\frac{\gamma^2+\gamma}{2})\right]e_p^2+\left[(2\gamma^2g_1g_2-\gamma^2(g_1^2+g_2^2)\right]e_{p1}e_p+ \\ (g_1-g_2)\gamma e_ye_{p1}-(g_1-g_2)\gamma e_ye_p \end{array}\right\} \quad (10)$$

Bias =B1 $(\hat{\bar{Y}}_7) = E\left(\hat{\bar{Y}}_7-\bar{Y}\right) =$

$$\bar{Y}\left\{\begin{array}{l} \left[g_2^2\left(\frac{\gamma^2+\gamma}{2}\right)+g_1^2\left(\frac{\gamma^2-\gamma}{2}\right)-\gamma^2g_1g_2\right]\lambda^1C_p^2+(2\gamma^2g_1g_2-\gamma^2(g_1^2+g_2^2))\lambda^1C_p^2+ \\ \left[g_1^2(\frac{\gamma^2-\gamma}{2})+g_2^2(\frac{\gamma^2+\gamma}{2})-\gamma^2g_1g_2\right]\lambda C_p^2+(g_1-g_2)\gamma\lambda^1\rho_{py}C_yC_p-(g_1-g_2)\gamma\lambda\rho_{yp}C_yC_p \end{array}\right\} \quad (11)$$

$$MSE_1(\hat{\bar{Y}}_7) = E\left(\hat{\bar{Y}}_7-\bar{Y}\right)^2 =$$

$$\bar{Y}^2E\left(\begin{array}{l} e_y^2+2(g_1-g_2)\gamma e_ye_{p1}-2(g_1-g_2)\gamma e_ye_p \\ -2(g_1-g_2)^2\gamma^2e_{p1}e_p+(g_1-g_2)^2\gamma^2e_{p1}^2+(g_1-g_2)^2\gamma^2e_p^2 \end{array}\right) \quad (12)$$

$$=\bar{Y}^2\left\{\begin{array}{l} \lambda C_y^2+2\left(\frac{1-a}{1+a}\right)\gamma\ \lambda^1\rho_{yp}C_yC_p-2\left(\frac{1-a}{a+1}\right)\gamma\lambda\rho_{yp}C_yC_p-2\left(\frac{1-a}{a+1}\right)^2\gamma^2\lambda^1C_p^2+ \\ \left(\frac{1-a}{1+a}\right)^2\gamma^2\lambda^1C_p^2+\left(\frac{1-a}{1+a}\right)^2\gamma^2\lambda C_p^2 \end{array}\right\}$$

$$MSE_1(\hat{\bar{Y}}_7) = \bar{Y}^2 \left\{ \lambda C_y^2 + 2\left(\frac{a-1}{1+a}\right)(\lambda - \lambda^1)\gamma\rho_{yp} C_y C_p + \left(\frac{a-1}{a+1}\right)^2 \gamma^2 (\lambda - \lambda^1) C_p^2 \right\} \quad (13)$$

To obtain the minimum mean square error, we differentiate (13) with respect to γ and set the resulting expression equal to zero. Thus,

$$MSE_{1\min}(\hat{\bar{Y}}_7) = \frac{\partial MSE(\hat{\bar{Y}}_7)}{\partial \gamma} = 0 \Rightarrow 2\left(\frac{a-1}{1+a}\right)(\lambda - \lambda^1)\rho_{yp} C_y C_p + 2\left(\frac{a-1}{1+a}\right)^2 \gamma(\lambda - \lambda^1) C_p^2 = 0$$

$$\Rightarrow \rho_{py} C_y C_p + \left(\frac{a-1}{1+a}\right)\gamma C_p^2 = 0 \Rightarrow k + \left(\frac{a-1}{1+a}\right)\gamma = 0 \Rightarrow \left(\frac{a-1}{1+a}\right)\gamma = -K \Rightarrow$$

$$(g_2 - g_1)\gamma = -K \Rightarrow \gamma = \left(\frac{K}{g_1 - g_2}\right) \Rightarrow \gamma = \frac{K(1+a)}{(1-a)}$$

$$\Rightarrow \gamma = \frac{K(1+a)}{(1-a)} \quad (14)$$

Where K= $\frac{\rho_{py} C_y}{C_p}$

Substituting (14) into (13), we have

$$MSE_{1\min}(\hat{\bar{Y}}_7) = \bar{Y}^2 C_y^2 \left\{ \lambda - \left(\lambda - \lambda^1\right)\rho_{yp}^2 \right\} \quad (15)$$

Where ρ_{yp} is the correlation coefficient between Y and P

Case II

Assuming the second phase sample size n^2 was drawn independently from the population, not the first phase sample n^1 then the suggested estimator would still be the same but the bias and MSE will be different in this case. The bias and MSE are obtained by taking $E(e_{x_1} e_{x_2}) = E(e_y e_{x_1}) = 0$

(16)

For case II of the proposed ratio-type estimator, the bias and mean square error are obtained by substituting (15) into (10) and (12).Thus , we have

$$Bias = B_2(\hat{\bar{Y}}_7) = E\left(\hat{\bar{Y}}_7 - \bar{Y}\right) =$$

$$\bar{Y}\left\{ \begin{array}{l} \left[g_2^2\left(\frac{\gamma^2+\gamma}{2}\right) + g_1^2\left(\frac{\gamma^2-\gamma}{2}\right) - \gamma^2 g_1 g_2 \right]\lambda^1 C_p^2 + \\ \left[g_1^2\left(\frac{\gamma^2-\gamma}{2}\right) + g_2^2\left(\frac{\gamma^2+\gamma}{2}\right) - \gamma^2 g_1 g_2 \right]\lambda C_p^2 - (g_1 - g_2)\gamma\lambda\rho_{yp} C_y C_p \end{array} \right\} \quad (17)$$

And mean square error

$$MSE_2(\hat{\bar{Y}}_7) = \bar{Y}^2\left\{\lambda C_y^2 - 2(g_1 - g_2)\gamma\lambda\rho_{yp}C_yC_p + (g_1 - g_2)^2\gamma^2\lambda^1 C_p^2 + (g_1 - g_2)^2\gamma^2\lambda\, C_p^2\right\}$$

$$= \bar{Y}^2\left\{\lambda C_y^2 - 2(g_1 - g_2)\gamma\lambda\rho_{yp}C_yC_p + (g_1 - g_2)^2\gamma^2(\lambda^1 + \lambda)C_p^2\right\}$$

$$MSE_2(\hat{\bar{Y}}_7) = \bar{Y}^2\left\{\lambda C_y^2 - 2\left(\frac{1-a}{1+a}\right)\gamma\lambda\rho_{yp}C_yC_p + \left(\frac{1-a}{1+a}\right)^2\gamma^2(\lambda^1 + \lambda)C_p^2\right\} \quad (18)$$

The minimum mean square error is obtained by differentiating (18) with respect to γ and equating to zero. Thus, we have

$$\frac{\partial MSE_2(\hat{\bar{Y}})}{\partial\gamma} = 0 \Rightarrow -2\left(\frac{1-a}{1+a}\right)\lambda\rho_{py}C_yC_p + 2\left(\frac{1-a}{1+a}\right)^2\gamma(\lambda + \lambda^1)C_P^2 = 0$$

$$\gamma = \frac{\lambda k}{\left(\frac{1-a}{1+a}\right)(\lambda + \lambda^1)} \quad (19)$$

Substituting (19) into (18) we have

$$MSE_{2\min}(\hat{\bar{Y}}_7) = \bar{Y}^2 C_y^2\left\{\lambda - w\rho_{yp}^2\right\} \quad (20)$$

Where $w = \frac{\lambda^2}{\lambda + \lambda^1}$

3. Efficiency Comparisons:

Bias, Relative bias, Mean Square Error, Relative Efficiency and Percentage Relative Efficiency

If $\hat{\bar{Y}}_7$ is the proposed estimator of the population mean, then the bias, relative bias, mean square error (MSE) and Relative Efficiency are given as:

(a) $$Bias = B(\hat{\bar{Y}}_7) = \left[E(\hat{\bar{Y}}_7) - \bar{Y}\right] \quad (21)$$

(b) $$\text{Relative bias} = RB(\hat{\bar{Y}}_7) = \left[\frac{E(\hat{\bar{Y}}) - \bar{Y}}{\bar{Y}}\right] \quad (22)$$

(c) Mean square error= $$MSE(\hat{\bar{Y}}_7) = E\left[(\hat{\bar{Y}}) - \bar{Y}\right]^2 \quad (23)$$

(d) Relative Efficiency: Let $MSE(\hat{\bar{Y}}_7)$ be the mean square error of the proposed estimator of the population mean and $MSE(\hat{\bar{Y}}_i)$ be the mean square error of any other estimator, then $MSE(\hat{\bar{Y}}_7)$ is said to be more efficient than $MSE(\hat{\bar{Y}}_i)$, if

(i) $$\frac{MSE(\hat{\bar{Y}}_7)}{MSE(\hat{\bar{Y}}_i)} < 1 \ or \ \frac{1}{\dfrac{MSE(\hat{\bar{Y}}_7)}{MSE(\hat{\bar{Y}}_i)}} > 1 \qquad (24)$$

(ii) $$MSE(\hat{\bar{Y}}_i) - MSE(\hat{\bar{Y}}_7) > 0 \qquad (25)$$

(e) Percentage Relative Efficiency (PRE): $MSE(\hat{\bar{Y}}_7)$ is said to be more efficient than $MSE(\hat{\bar{Y}}_i)$ in terms of PRE if

$$\frac{MSE(\hat{\bar{Y}}_7)}{MSE(\hat{\bar{Y}}_i)} X100 < 100 \ or \ \frac{1}{\dfrac{MSE(\hat{\bar{Y}}_7)}{MSE(\hat{\bar{Y}}_i)}} X100 > 100 \qquad (26)$$

Note : An estimator is said to be more efficient than its counterpart estimator if the estimator is one with a minimum mean square error. Also, an estimator is said to be percentage relative efficient in relation to another estimator, if it's one with the largest Percentage relative efficiency.

4. Consider the following estimators

(i) Sample mean (SRSWOR). From (1) and (15) we have

$$MSE(\hat{\bar{Y}}_1) - MSE_{\min}(\hat{\bar{Y}}_7) \Rightarrow \bar{Y}^2 C_y^2 \left(\lambda - \lambda^1\right)\rho_{py}^2 \geq 0$$

Whenever $\rho_{yp}^2 \geq 0$ (since $\bar{Y}^2 > 0, C_y^2 > 0$ and $\left(\lambda - \lambda^1\right) = \left(\frac{1}{n} - \frac{1}{n^1}\right) > 0$) (27)

(ii) Double sampling ratio estimator using auxiliary attribute. From (3) and (15) we have

$$MSE(\hat{\bar{Y}}_3) - MSE_{\min}(\hat{\bar{Y}}_7) = \bar{Y}^2\left[\lambda C_y^2 + \left(\lambda - \lambda^1\right)\left(C_p^2 - 2\rho_{py}C_yC_p\right)\right] - \bar{Y}^2 C_y^2\left[\lambda - (\lambda - \lambda^1)\rho_{py}^2\right]$$

$$= \left(\lambda - \lambda^1\right)\bar{Y}^2\left(C_p - \rho_{py}C_y\right)^2 \geq 0$$

Since $\left(C_p - \rho_{py}C_y\right)^2 \geq 0$, $\left(\lambda - \lambda^1\right) \geq 0$ and $\bar{Y}^2 \geq 0$. (28)

(iii) Double sampling exponential ratio estimator using auxiliary attribute. From (5) and (15), we have

$$MSE(\hat{\bar{Y}}_5) - MSE(\hat{\bar{Y}}_7) = \bar{Y}^2\left[\lambda C_y^2 + (\lambda - \lambda^1)\left(\frac{1}{4}C_p^2 - \rho_{py}C_yC_p\right)\right] - \bar{Y}^2 C_y^2\left[\lambda - (\lambda - \lambda^1)\rho_{py}^2\right]$$

$$= \bar{Y}^2(\lambda - \lambda^1)\left[\rho_{py}C_y - \frac{C_p}{2}\right]^2 \geq 0$$

Since $\bar{Y}^2 \geq 0,\ (\lambda - \lambda^1) \geq 0\ and\ \left(\rho_{py}C_y - \frac{C_p}{2}\right)^2 \geq 0$ (29)

(iv) Double sampling generalized family of ratio estimator using auxiliary variable. From (6) and (15) we have

$$MSE(\hat{\bar{Y}}_6) - MSE(\hat{\bar{Y}}_7) = \bar{Y}^2 C_y^2\left(\lambda - (\lambda - \lambda^1)\rho_{px}^2\right) - \bar{Y}^2 C_y^2\left(\lambda - (\lambda - \lambda^1)\rho_{py}^2\right)$$

$$= \bar{Y}^2 C_y^2(\lambda - \lambda^1)\left(\rho_{py}^2 - \rho_{yx}^2\right) \geq 0$$

If $\rho_{py}^2 \geq \rho_{yx}^2$ (30)

5. Empirical Studies

In this section, various results obtained in the previous sections are now examined with the help of the following data:

Population I: (Source: William G. Cochran (1977), page 34)

Y= Food Cost

X= Family Income

$\Theta =$ Family of size more than 3

N=33	n^2=16	n^1=22	$\bar{Y} = 1.102$	C_x=0.146
C_y=0.369	C_p=2.7005	P=0.124	$\rho_{py} = 0.49987$	$\rho_{yx} = 0.2521$

Population II: (Source: Advance Data from vital and Health statistics, number 347, October, 2004(CDC))

Y= Height of the people

X= Weight of the people

$\Theta =$ Sex of the people

N=36	n^2=12	n^1=22	$\bar{Y} = 140.18$	$C_x = 0.482337$
$C_x = 0.191654$	$C_p = 1.014$	P=0.50	$\rho_{py} = 0.963$	$\rho_{yx} = 0.973$

Table 1 MSE of various estimators

ESTIMATOR	MSE	
	POPULATION	
	I	II
$\hat{\bar{Y}}_1$	3.7172	40.0992
$\hat{\bar{Y}}_2$	3.6002	79.36818
$\hat{\bar{Y}}_3$	17.4947	526.8215
$\hat{\bar{Y}}_4$	3.5520	16.4415
$\hat{\bar{Y}}_5$	5.7500	92.1299
$\hat{\bar{Y}}_6$ (case I)	3.5440	14.2153
$\hat{\bar{Y}}_6$ (case II)	3.5307	11.2996
$\hat{\bar{Y}}_7$ (case I)	3.3069	14.7446
$\hat{\bar{Y}}_7$ (case II)	3.2754	11.8885

Table 2 Pre of the Various Estimators with Respect to the Sample Mean

ESTIMATORS	PRE	
	POPULATION	
	I	II
$\hat{\bar{Y}}_1$	100	100
$\hat{\bar{Y}}_2$	103.25	50.52
$\hat{\bar{Y}}_3$	21.25	7.61
$\hat{\bar{Y}}_4$	104.65	243.89
$\hat{\bar{Y}}_5$	64.65	43.52
$\hat{\bar{Y}}_6$ (case I)	104.89	282.08
$\hat{\bar{Y}}_6$ (case II)	105.28	354.87

$\hat{\bar{Y}}_7$ (case I)	112.41	271.96
$\hat{\bar{Y}}_7$ (case II)	113.49	337.29

6. Conclusion

From table 1 and 2, we can see that under population I, the proposed estimator ($\hat{\bar{Y}}_7$) under double sampling using auxiliary attribute is more efficient in terms of MSE and PRE than the existing sample mean $\hat{\bar{Y}}_1$, the double sampling ratio estimator using auxiliary variable $\hat{\bar{Y}}_2$, the double sampling ratio estimator using auxiliary attribute $\hat{\bar{Y}}_3$, the double sampling exponential ratio estimator using auxiliary variable $\hat{\bar{Y}}_4$, the double sampling exponential ratio estimator using auxiliary attribute $\hat{\bar{Y}}_5$ and the generalized family of ratio estimator $\hat{\bar{Y}}_6$ with auxiliary variable. Also, from population II, we can see that the generalized family of ratio estimator using auxiliary variable is more efficient than the proposed estimator as a result of the correlation between the study variable y and the auxiliary attribute Θ.

i.e $\rho_{yx} > \rho_{py} \Rightarrow$ 0.973>0.963, which is contrary to the condition for its efficient performance as stated in (30). For both the existing generalized family of ratio estimator and the proposed estimator, we can see that case II is more efficient than case I for both populations

Thus, we can conclude that using auxiliary attribute with high correlation with the study variable can boost variable estimation. Also, using double sampling scheme with the second phase sample size selected independently from the population instead of the first phase sample is more efficient.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Bahl, S. and Tuteja. R.K. (1991). Ratio and Product type Exponential Estimator. Information and Optimization sciences, 12(1), 159-163.

[2] Bhusan, S. (2013). Improved sampling strategies in finite population. Scholars Press, Germany.

[3] Bhushan et. al. (2017). On the class of Double Sampling Ratio Estimators using Auxiliary Information on an Attribute and on Auxiliary variable. International Journal of Statistics and Systems, 12(1), 15-23

[4] Cochran, G.O (1977). Sampling Techniques, 3rd ed. New York, John Wiley and Sons

[5] Etorti et al. (2023). A Generalized Efficient Family of Ratio Estimator of Population mean under Double Sampling Techniques. International Research Journal of Innovations in Engineering and Technology, 7(12) 189-202

[6] Jhajj, H.S, Sharma, M.K and Grover, L.K (2006): A family of Estimators of Population Mean using the Information on an Auxiliary Attribute. Pak. J. Statist 22(1), 43-50,

[7] Kadilar G.O (2016). A New Exponential Type Estimator for the population mean in Simple Random Sampling. Journal of Modern Applied Statistical Methods, 15, 207-214

[8] Naik, V.D and Gupta, P.C (1996): A note on Estimation of mean with known Population Proportion of an Auxiliary Character. Jour. Ind. Soc. Agri. Stat. 48(2), 151-158

[9] Rather et. al. (2022). A New Exponential Ratio Type Estimator for the Population Means using Information on Auxiliary Attribute. Journal of Statistics Applications & Probability Letters, 9(1),31-42

[10] Scrivastava, S.K (1967). An Estimator using Auxiliary Information. Calcutta Statistical Association Bulletin, 16, 121-132

[11] Singh et. al. (2003). On Linear Regression and Ratio-product Estimation of a Finite Population Mean. The Statistician, 52 (1), 59-67

[12] Shabbir, J. and Gupta, S. (2007): On Estimating the Finite population mean with known population proportion of an auxiliary variable. Pak. J. Statist.,23(1), 1-9