A Class of Exponential Estimator to Estimate the Population Mean in the Presence of Non-Response

Napattchan Dansawad

Department of Applied Mathematics, Faculty of Science and Technology, Valaya Alongkorn Rajabhat University under the Royal Patronage, Phathum Thani, 13180, Thailand

Corresponding author. E-mail address: napattchan@vru.ac.th

Received: 6 March 2019; Revised: 7 June 2019; Accepted: 10 June 2019

Abstract

This paper presents a class of exponential estimator to estimate the population mean in the presence of non-response using a small amount of known auxiliary information, when non-response occurs only on the study variable. Furthermore, the study also proposes a modified exponential estimator using a linear combination of a few members of the proposed class. Properties of the proposed estimators are obtained up to first order approximation. Results from theoretical and empirical studies have shown that the proposed modified exponential estimator at its optimum performs more efficient than all other relevant estimators.

Keywords: Exponential estimator, Study variable, Auxiliary variable, Non-response

Introduction

In sample surveys, the auxiliary information has been efficiently used by several statisticians to improve the precision of their estimates for the population mean. As an example, Sisodia and Dwivedi (1981), Upadhyaya and Singh (1984), Singh and Kakran (1993). Singh and Tailor (2003), Kadilar and Cingi (2006a, 2006b), Khoshnevisan, Singh, Chauhan, Sawan, and Smarandache (2007), and Rachokarn and Lawson (2016) demonstrated the use of auxiliary information in defining the modified ratio estimators under the simple random sampling scheme.

Besides the use of auxiliary information, several other statisticians including Singh et al. (2009) have used the exponential type estimator along with the auxiliary information to estimate population mean by extending the work of Kadilar and Cingi (2006b) and Khoshnevisan et al. (2007) and propose an exponential family of estimator for the population mean in the simple random sampling (T_1), as

$$T_1 = \bar{y}\exp\left(\frac{(a\bar{X}+b)-(a\bar{x}+b)}{(a\bar{X}+b)+(a\bar{x}+b)}\right) \tag{1}$$

Where $\bar{y}$ and $\bar{x}$ are the sample means of study and auxiliary variables respectively. While $\bar{X}$ is the population mean of auxiliary variable, whereas $a(\neq 0)$, and b are either real numbers or the functions of known parameters of auxiliary variable.

However, one of the most important problems of sample survey is non-response. It is the phenomenon that happens when the required information is not obtained from the persons selected in the sample. An estimate obtained from incomplete required information can lead to misleading interpretation and conclusions. In order to reduce the effect of non-response in such situations, Hansen and Hurwitz (1964) gave the technique of sub sampling from non- respondents in the sample and proposed an unbiased estimator of population mean by combining the information available from response and non-response groups.

Assume that the population of size N on the study variable y and the auxiliary variable x is supposed to be divided into two groups, N_1 response group and $N_2 = N - N_1$ non-response group respectively. Let a sample of size n be drawn from the population of size N by using simple random sampling without replacement (SRSWOR), it has been observed that n_1 units respond and n_2 units do not respond. From the n_2 non-response units, a sub-sample of size $r = n_2 / k, k > 1$ is drawn by making extra efforts. Therefore, the unbiased estimator for $\bar{Y}$ based on $n_1 + r$ units on study variable y is defined by Hansen and Hurwitz (1964) as follows:

$$\bar{y}^* = w_1\bar{y}_1 + w_2\bar{y}_{2r} \tag{2}$$

Where $w_1 = n_1 / n,\ w_2 = n_2 / n,\ \bar{y}_1,$ and $\bar{y}_{2r}$ are the sample means of study variable y based on n_1 and r units respectively. The variance of $\bar{y}^*$ is given by:

$$V(\bar{y}^*) = \lambda S_y^2 + \lambda^* S_{y(2)}^2 \tag{3}$$

Where $\lambda = (N-n)/Nn,\quad \lambda^* = W_2(k-1)/n,$ and $W_2 = N_2 / N.\quad S_y^2 = \sum_{i=1}^{N}(y_i - \bar{Y})^2/(N-1)$ and $S_{y(2)}^2 = \sum_{i=1}^{N_2}(y_i - \bar{Y}_2)^2/(N_2 - 1)$ are the variances for the entire population and the non- response group, respectively. While $\bar{Y}_2 = \sum_{i=1}^{N_2} y_i / N_2$ is the population mean of the non-response group.

The objective of this paper is to extend the work of Singh et al. (2009), by proposing a class of exponential estimator to estimate the population mean in the presence of non- response when non- response occurs only on the study variable. In addition, this study also presents a modified exponential estimator by combining several previous proposed estimators and compares the efficiency of the proposed modified exponential estimator with respect to other relevant estimators. Properties of the proposed estimators, especially the mean squared errors of these estimators have been obtained up to the first order of approximation. The theoretical results are supported by an empirical study.

The Proposed Class of Estimator

Motivated by Singh et al. (2009), a class of exponential estimator to estimate the population mean $\bar{Y}$ of the study variable y when non-response occurs only on the study variable as below was investigated.

$$T^* = \bar{y}^* \exp\left(\frac{(a\bar{X}+b)-(a\bar{x}+b)}{(a\bar{X}+b)+(a\bar{x}+b)}\right) \tag{4}$$

Where T^* is the proposed class of exponential estimator, $a(\neq 0),\ b$ are either real constants or the functions of the known parameters such as the standard deviation (S_x), coefficient of variation (C_x), kurtosis $(\beta_2(x))$ and correlation coefficient (ρ_{yx}) of auxiliary variable x.

To obtain the properties of the estimator such as bias and MSE of the proposed class of exponential estimator T^*, define the following relative error terms and their expectations as $\bar{y}^* = \bar{Y}(1+e_0^*),\ \bar{x} = \bar{X}(1+e_1)$ such that $E(e_0^*) = E(e_1) = 0,\ E(e_0^{*2}) = \lambda C_y^2 + \lambda^* C_{y(2)}^2,\ E(e_1^2) = \lambda C_x^2,\ E(e_0^*, e_1) = \lambda C C_x^2,$

Where $C_y^2 = S_y^2 / \bar{Y}^2$, $C_{y(2)}^2 = S_{y(2)}^2 / \bar{Y}^2$, $C_x^2 = S_x^2 / \bar{X}^2$, $C = \rho_{yx} C_y / C_x$.

Rewriting equation (4) in terms of e's, turns into the below equation

$$T^* = \bar{Y}(1+e_0^*)\exp\{-\tau e_1(1+\tau e_1)^{-1}\} \tag{5}$$

Where $\tau = a\bar{X}/(a\bar{X}+b)$.

Expanding the right hand side of equation (5), multiplying out and neglecting terms involving powers of e's greater than two, as below

$$T^* \approx \bar{Y}\left(1+e_0^* - \tau e_1 - \tau e_0^* e_1 + \tau^2 e_1^2\right) \tag{6}$$

Subtracting $\bar{Y}$ from both sides of equation (6), and taking expectations of both sides, one gains the bias of the proposed class of exponential estimator T^* to the first order of approximation, as

$$Bias(T^*) = \lambda \bar{Y} C_x^2 \tau(\tau - C) \tag{7}$$

Subtracting $\bar{Y}$ and squaring both sides of equation (6), and then by taking expectations of both sides, one gains the MSE of the proposed class of exponential estimator T^* to the first order of approximation, as

$$MSE(T^*) = \bar{Y}^2\left[\lambda\left[C_y^2 + \tau C_x^2(\tau - 2C)\right] + \lambda^* C_{y(2)}^2\right] \tag{8}$$

A Few Members of the Proposed Class of Exponential Estimator

A few of the existing estimators listed in Table 1 are members of the following proposed class of exponential estimator T^* mentioned in equation (4) which can be obtained by assigning suitable the different values of constants a and b.

Table 1 A few members of the proposed class of exponential estimator T^*

Estimator	Values of Constants	
	a	b
$T_1^* = \bar{y}^* \exp\left(\dfrac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right)$	1	0
$T_2^* = \bar{y}^* \exp\left(\dfrac{(\bar{X}-\bar{x})}{(\bar{X}+\bar{x})+2C_x}\right)$	1	C_x
$T_3^* = \bar{y}^* \exp\left(\dfrac{(\bar{X}-\bar{x})}{(\bar{X}+\bar{x})+2\beta_2(x)}\right)$	1	$\beta_2(x)$
$T_4^* = \bar{y}^* \exp\left(\dfrac{(\bar{X}-\bar{x})}{(\bar{X}+\bar{x})+2\rho_{yx}}\right)$	1	ρ_{yx}
$T_5^* = \bar{y}^* \exp\left(\dfrac{\beta_2(x)(\bar{X}-\bar{x})}{\beta_2(x)(\bar{X}+\bar{x})+2C_x}\right)$	$\beta_2(x)$	C_x

Table 1 (Cont).

Estimator	Values of Constants	
	a	b
$T_6^* = \bar{y}^* \exp\left(\dfrac{C_x(\bar{X}-\bar{x})}{C_x(\bar{X}+\bar{x})+2\beta_2(x)}\right)$	C_x	$\beta_2(x)$
$T_7^* = \bar{y}^* \exp\left(\dfrac{C_x(\bar{X}-\bar{x})}{C_x(\bar{X}+\bar{x})+2\rho_{yx}}\right)$	C_x	ρ_{yx}
$T_8^* = \bar{y}^* \exp\left(\dfrac{\rho_{yx}(\bar{X}-\bar{x})}{\rho_{yx}(\bar{X}+\bar{x})+2C_x}\right)$	ρ_{yx}	C_x
$T_9^* = \bar{y}^* \exp\left(\dfrac{\beta_2(\bar{X}-\bar{x})}{\beta_2(\bar{X}+\bar{x})+2\rho_{yx}}\right)$	$\beta_2(x)$	ρ_{yx}
$T_{10}^* = \bar{y}^* \exp\left(\dfrac{\rho_{yx}(\bar{X}-\bar{x})}{\rho_{yx}(\bar{X}+\bar{x})+2\beta_2(x)}\right)$	ρ_{yx}	$\beta_2(x)$

On substituting the values of a and b from Table 1 in equation (8), the MSEs of various members of the proposed class of exponential estimator T^* to the first degree of approximation, are obtained as

$$MSE(T_i^*) \approx \bar{Y}^2\left[\lambda\left[C_y^2+\tau_i C_x^2(\tau_i-2C)\right]+\lambda^* C_{y(2)}^2\right]; i=1,...,10. \qquad (9)$$

Where $\tau_1 = 1$, $\tau_2 = \dfrac{\bar{X}}{(\bar{X}+C_x)}$, $\tau_3 = \dfrac{\bar{X}}{(\bar{X}+\beta_2(x))}$, $\tau_4 = \dfrac{\bar{X}}{(\bar{X}+\rho_{yx})}$, $\tau_5 = \dfrac{\beta_2(x)\bar{X}}{(\beta_2(x)\bar{X}+C_x)}$, $\tau_6 = \dfrac{C_x\bar{X}}{(C_x\bar{X}+\beta_2(x))}$,

$\tau_7 = \dfrac{C_x\bar{X}}{(C_x\bar{X}+\rho_{yx})}$, $\tau_8 = \dfrac{\rho_{yx}\bar{X}}{(\rho_{yx}\bar{X}+C_x)}$, $\tau_9 = \dfrac{\beta_2(x)\bar{X}}{(\beta_2(x)\bar{X}+\rho_{yx})}$, $\tau_{10} = \dfrac{\rho_{yx}\bar{X}}{(\rho_{yx}\bar{X}+\beta_2(x))}$.

Modified Exponential Estimator

In this section, the proposed class of exponential estimator T^* in the section 2 can be extended as a modified exponential estimator by forming linear combination of T_1^* and $T_i^*; (i=1,...,10)$ estimators, as follows:

$$T_i^{**} = \omega T_1^* + (1-\omega)T_i^* \quad ; i=1,...,10 \qquad (10)$$

Where T_i^{**} is the modified exponential estimator, whereas ω is a suitable choice of constant which make the MSE of T_i^{**} minimum and $T_i^*; (i=1,...,10)$ are estimators listed in Table 1.

Now consider adopting the same procedure in section 2, one can obtain the MSE of T_i^{**}, to the first degree of approximation, as

$$MSE(T_i^{**}) = \bar{Y}^2\left[\lambda\left[C_y^2+\left(\frac{\omega}{2}+\tau-\omega\tau\right)C_x^2\left(\frac{\omega}{2}+\tau-\omega\tau-2C\right)\right]+\lambda^* C_{y(2)}^2\right] \qquad (11)$$

The MSE of T_i^{**} is minimized for

$$\omega = \frac{2\left[C-\tau_i(1+2C-2\tau_i)\right]}{\left[1+4\tau_i(\tau_i-1)\right]} = \omega_{opt.} \qquad (12)$$

Substituting (12) in (11), the optimum MSE of T_i^{**} is given by

$$MSE(T_{i(opt.)}^{**}) = \bar{Y}^2\left[\lambda C_y^2(1-\rho_{yx}^2)+\lambda^* C_{y(2)}^2\right] \qquad (13)$$

Efficiency Comparisons

In this section, to compare the efficiency of the proposed class of exponential estimator T^* in (4) with the unbiased estimator $\bar{y}^*$ in (2), one can obtain after some algebra the following equation:

$$V(\bar{y}^*) - MSE(T^*) = (\tau - 2C) > 0 \qquad (14)$$

When condition (14) is satisfied, one can conclude that the proposed class of exponential estimator T^* outperforms the unbiased estimator $\bar{y}^*$.

Next, the efficiency of the proposed modified exponential estimator $T_i^{**};\ (i=1,...,10)$ in (10) is compared with the proposed class of exponential estimator listed in Table 1. The following condition was used:

$$MSE(T_i^*) - MSE(T_{i(opt.)}^{**}) = (\tau_i - C)^2 > 0 \qquad i=1,...,10 \qquad (15)$$

If this condition (15) is satisfied, one can infer that all proposed modified exponential estimator $T_i^{**};\ (i=1,...,10)$ are more efficient than the proposed class of exponential estimator $T_i^*;\ (i=1,...,10)$.

Empirical study

To illustrate the performance of various estimators of population mean $\bar{Y}$, one can consider a real data set used before by Khare and Sinha (2004). The data belongs to the population census of 96 villages in a rural area published by the Government of India for the West Bengal state in 1981. This study assumed that the number of agricultural labors in the village was taken as study variable y and the area of the village as taken as auxiliary variable x. In addition, the number of villages whose area was greater than 160 hectares was considered as the non-response group of the population which equal 24 villages. The descriptions of the parameters related to the study variable y and the auxiliary variable x under study have been given in Table 2 below,

Table 2 Parameters and constants of the population under study

$N=96$	$\bar{X}=144.87$	$\bar{Y}=137.93$
n=40	$C_x=0.81$	$C_y=1.32$
$\rho_{yx}=0.77$	$C_{x(2)}=0.94$	$C_{y(2)}=2.08$

For the purpose of efficiency comparison of the different estimators listed in Table 1, all estimators with respect to the unbiased estimator $\bar{y}^*$ were compared using the percent relative efficiencies (PREs) criterion for the varying values of k. The results are compiled in Table 3.

Table 3 Percent relative efficiencies (PREs) of the proposed estimator with respect to $\bar{y}^*$ for the different values of k

Estimator	$1/k$			
	1/5	1/4	1/3	1/2
$\bar{y}^*$	100.0000	100.0000	100.0000	100.0000
T_1^*	107.7573	109.9211	113.7589	122.4393
T_2^*	107.7654	109.9317	113.7742	122.4661
T_3^*	107.6442	109.7736	113.5473	122.0683
T_4^*	107.7671	109.9339	113.7773	122.4716
T_5^*	107.7910	109.9651	113.8222	122.5503
T_6^*	104.6315	105.8758	108.0345	112.7002
T_7^*	107.7593	109.9237	113.7628	122.4460
T_8^*	107.7551	109.9182	113.7549	122.4322
T_9^*	107.7915	109.9657	113.8230	122.5518
T_{10}^*	107.5993	109.7150	113.4634	121.9214
$T_{(opt.)}^{**}$	112.8378	116.6376	123.6323	140.7742

Table 3 clearly indicates that the proposed estimator is more efficient than the usual unbiased estimator $\bar{y}^*$. The PREs of all estimators decrease as the value of k increases. Also, the performance of the proposed estimator $T_i^*;(i=2,...,10)$ at its optimum conditions is the best among all other estimators proposed and listed in Table 1.

Conclusion

The authors of this research have extended the work carried out by Singh et al. (2009) by proposing a class of exponential estimator to estimate the population mean in the presence of non-response when members of the proposed class are also derived by allocating the different values of constants, listed in Table 1. In addition, a modified exponential estimator was developed by forming linear combination between a few members of the proposed class of exponential estimator in the previous context. The theoretic and empirical study proved that the proposed modified exponential estimator at its optimum performs better than the usual unbiased estimator $\bar{y}^*$ and all other relevant estimators listed in Table1.

References

Hansen, M. H., & Hurwitz, W. N. (1946). The problem of non-response in sample surveys, *Journal of the American Statistical Association, 41,* 517-529.

Kadilar, C., & Cingi, H. (2006a). A new ratio estimator using correlation coefficient. *Statistics on the Internet, 12,* 1-11.

Kadilar, C., & Cingi, H. (2006b). Improvement in estimating the population mean in simple random sampling. *Applied Mathematics Letters, 19,* 75-79.

Khare, B. B., & Sinha, R. R. (2004). Estimation of finite population ratio using two-phase sampling scheme in the presence of non-response. *Aligarh Journal of Statistics, 24,* 43-56.

Khoshnevisan, M., Singh, R., Chauhan, P., Sawan, N., & Smarandache, F. (2007). A general family of estimators for estimating population mean using known value of some population parameter(s). *Far East Journal of Theoretical Statistics, 22,* 181-191.

Rachokarn, T., & Lawson, N. (2016). An efficient family of estimators for the population mean using auxiliary information in the presence of missing observations. *Proceedings of the 2nd international conference on mathematics, engineering, and industrial application 2016,* 10-12 August 2016 (pp. 1-6). Songkhla, Thailand: AIP Conference Proceedings.

Singh, H. P., & Kakran, M. S. (1993). A modified ratio estimator using coefficient of variation of auxiliary character. N.P.: n.p.

Singh, R., Chauhan, P., Sawan, N., & Smarandache, F. (2009). Improvement in estimating the population mean using exponential estimator in simple random sampling. *Bulletin of Statistics and Economics, 3,* 13-18.

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.

Upadhyaya, L. N., & Singh, H. P. (1984). On the estimation of population mean with known coefficient of variation. *Biometrical Journal, 26,* 915-922.