STATISTICS IN TRANSITION new series, March 2018
Vol. 19, No. 1, pp. 159–166, DOI 10.21307/stattrans-2018-010

PRODUCT EXPONENTIAL METHOD OF IMPUTATION IN SAMPLE SURVEYS

Shakti Prasad[1]

ABSTRACT

In this paper, a product exponential method of imputation has been suggested and their corresponding resultant point estimator has been proposed for estimating the population mean in sample surveys. The expression of bias and the mean square error of the suggested estimator has also been derived, up to the first order of large sample approximations. Compared with the mean imputation method, Singh and Deo (Statistical Papers (2003)) and Adapted estimator (Bahl and Tuteja (1991)), the simulation studies show that the suggested estimator is the most efficient estimator.

Key words: imputation methods, bias, mean square error (MSE), efficiency.

1. Introduction

The use of auxiliary information for estimating the finite population mean of the study variable has played an eminent role in sample surveys. The ratio imputation method is employed for missing data if the correlation between the study variable and the auxiliary variable is positive. On the other hand, if this correlation is negative, the product imputation method investigated by Singh and Deo (2003), is quite effective. The application of the product imputation method has too much importance but absolutely has some limitation in medical discipline, industrial and social science, etc. There are several medical or social science related variables which decrease as the people grow up. For example, as the people become older, the following variables have negative correlation with the age: (a) duration of sleeping hours (b) hearing tendency (c) eye sight (d) number of hairs on the head (e) number of love affairs and (f) working hour capacity etc. If information on any of these study variable is missing, but the age of the persons is available, the product imputation method will be beneficial.

It is worth to be noticed that appreciable amount of works carried out under product method of estimation in sample surveys by several authors, but its application very limited in imputation methods. Dated back, Singh and Deo (2003) have used the product imputation method in survey sampling. Motivated with the above work, we study the some product exponential method of imputation in sample surveys.

Let y and x be denoted by the negatively correlated study variable and auxiliary variable respectively. A simple random sample (without replacement) s_n of

[1]Department of Basic & Applied Science National Institute of Technology, Arunachal Pradesh,Yupia, Papum pare-791112, India. E-mail: shakti.pd@gmail.com.

n units is depleted from a finite population $U = (U_1, U_2, ..., U_N)$ of N units to estimate the population mean $\bar{Y}$. Let r be the number of responding units out of sampled n units, the set of responding units by R and the set of non-responding units by R^c. If the units involve the responding unit set, the values on the study variable y_i are observed. If they involve the non-responding unit set, the values on the study variable y_i are missing and hereafter the imputed values are derived for a well known units.

$$y_{.i} = \begin{cases} y_i & \text{if } i\varepsilon R \\ \tilde{y}_i & \text{if } i\varepsilon R^c \end{cases} \tag{1}$$

The general point estimator of population mean $\bar{Y}$ takes the form:

$$\bar{y}_s = \frac{1}{n}\sum_{i\varepsilon s_n} y_{.i} = \frac{1}{n}\left[\sum_{i\varepsilon R} y_{.i} + \sum_{i\varepsilon R^c} y_{.i}\right] = \frac{1}{n}\left[\sum_{i\varepsilon R} y_i + \sum_{i\varepsilon R^c} \tilde{y}_i\right] \tag{2}$$

Here, the value $\tilde{y}_i$ denote the imputed value of the study variable corresponding to the i^{th} non- responding units.

The consequently notations have been approaching in this work:

$\bar{Y}, \bar{X}$: The population means of the variables y and x respectively.

$\bar{y}_r, \bar{x}_r$: The response means of the respective variables for the sample sizes shown in suffices.

$\bar{x}_n$: The sample mean of the variable x.

ρ_{yx} : The population correlation coefficient between the variables y and x.

$S_x^2 = (N-1)^{-1}\sum_{i=1}^{N}(x_i - \bar{X})^2$: The population mean square of the variable x.

S_y^2 : The population mean square of the variable y .

C_y and C_x : The coefficients of variation of the variables shown in suffices.

2. Some Existing Estimators

In this section, the several estimators with imputation have been discussed for estimating the population mean in sample surveys.

2.1. Mean Imputation Method

Under this method, After imputation, data take the form:

$$y_{.i} = \begin{cases} y_i & \text{if } i\varepsilon R \\ \bar{y}_r & \text{if } i\varepsilon R^c \end{cases} \tag{3}$$

The resultant point estimator (2) of $\bar{Y}$ becomes

$$\bar{y}_m = \frac{1}{r}\sum_{i=1}^{r} y_i = \bar{y}_r \tag{4}$$

which is known as the response mean estimator $\bar{y}_r$ of population mean $\bar{Y}$.
The variance of the response sample mean $\bar{y}_r$, is given by

$$Var(\bar{y}_r) = \left(\frac{1}{r} - \frac{1}{N}\right)\bar{Y}^2 C_y^2 \tag{5}$$

2.2. Product Imputation Method

Singh and Deo (2003) proposed the product imputation method in sample surveys. After imputation, data take the form:

$$y_{.i} = \begin{cases} y_i & \text{if } i\varepsilon R \\ \bar{y}_r \left[\frac{n\bar{x}_r - r\bar{x}_n}{\bar{x}_n}\right] \frac{x_i}{\sum_{i\varepsilon R^c} x_i} & \text{if } i\varepsilon R^c \end{cases} \tag{6}$$

Under this method of imputation, the resultant point estimator (2) of $\bar{Y}$ becomes

$$\bar{y}_{SD} = \bar{y}_r \frac{\bar{x}_r}{\bar{x}_n} \tag{7}$$

which is analogue of the product estimator proposed by Murthy (1964).
The MSE of estimator $\bar{y}_{SD}$, is given by

$$MSE(\bar{y}_{SD}) = Var(\bar{y}_r) + \left(\frac{1}{r} - \frac{1}{n}\right)\bar{Y}^2\left(C_x^2 + 2\rho_{yx}C_yC_x\right) \tag{8}$$

3. Adapted Product Exponential Method of Imputation

Following the Bahl and Tuteja (1991), We have adapted product exponential method of imputation and their corresponding estimator for estimating the $\bar{Y}$ in survey sampling.
The adapted imputation method,
After imputation, the data take the form:

$$y_{.i} = \begin{cases} y_i & \text{if } i\varepsilon R \\ \frac{\bar{y}_r}{n-r}\left[n\exp\left(\frac{\bar{x}_r - \bar{X}}{\bar{x}_r + \bar{X}}\right) - r\right] & \text{if } i\varepsilon R^c \end{cases} \tag{9}$$

Under above adapted imputation methods, the resultant point estimators (2) of the population mean $\bar{Y}$ become

$$\bar{y}_{AE} = \bar{y}_r \exp\left[\frac{\bar{x}_r - \bar{X}}{\bar{x}_r + \bar{X}}\right] \tag{10}$$

The MSE of estimator $\bar{y}_{AE}$, is given by

$$MSE(\bar{y}_{AE}) = \left(\frac{1}{r} - \frac{1}{N}\right)\left(C_y^2 + \frac{1}{4}C_x^2 + \rho_{yx}C_yC_x\right)\bar{Y}^2 \quad (11)$$

4. Suggested Method and their Estimator

Following the Prasad (2016 & 2017), a product exponential method of imputation and their corresponding estimator has suggested for estimating the population mean $\bar{Y}$ in sample surveys.
The suggested imputation method,
After imputation, the data take the form:

$$y_{.i} = \begin{cases} \phi y_i \exp\left(\frac{(\bar{x}_r - \bar{X})S_x}{(\bar{X}+\bar{x}_r)S_x + 2C_x}\right) & \text{if } i\varepsilon R \\ \phi \frac{\bar{y}_r}{\bar{x}_r}\left(x_i - \frac{n}{n-r}(\bar{x}_n - \bar{x}_r)\right)\exp\left(\frac{(\bar{x}_r - \bar{X})S_x}{(\bar{X}+\bar{x}_r)S_x + 2C_x}\right) & \text{if } i\varepsilon R^c \end{cases} \quad (12)$$

Under above suggested imputation method, the resultant point estimator (2) of the population mean $\bar{Y}$ becomes

$$\zeta = \phi \bar{y}_r \exp\left[\frac{(\bar{x}_r - \bar{X})S_x}{(\bar{X}+\bar{x}_r)S_x + 2C_x}\right] \quad (13)$$

where ϕ is suitably chosen constant, such that the MSE of the resultant point estimator is minimum. It has been assumed that S_x and C_x are known.

5. Properties of the suggested estimator ζ

The bias and their mean square error (MSE) of the suggested estimator ζ are derived up to the first order of large sample approximations under the following transformations:
$\bar{y}_r = \bar{Y}(1+e_y)$ and $\bar{x}_r = \bar{X}(1+e_x)$ such that $E(e_i) = 0$, $|e_i| < 1 \forall\ i = y, x$.
Using the above transformations, the estimator ζ take the following form:

$$\zeta = \phi \bar{Y}(1+e_y)\, exp\left[\frac{1}{2}\theta e_x\left(1 + \frac{1}{2}\theta e_x\right)^{-1}\right] \quad (14)$$

where $\theta = \frac{\bar{X}S_x}{\bar{X}S_x + C_x}$.
Neglecting the higher power terms of $e's$, the equation(14) can be written as

$$\zeta - \bar{Y} \cong \bar{Y}\left[(\phi - 1) + \phi\left(e_y + \frac{1}{2}\theta e_x + \frac{1}{2}\theta e_y e_x - \frac{1}{8}\theta^2 e_x^2\right)\right] \quad (15)$$

Taking expectation of (15), we obtained the bias of the suggested estimator, is given as

$$Bias(\zeta) = \bar{Y}\left[(\phi-1) - \frac{1}{8}\theta\phi\left(\frac{1}{r}-\frac{1}{N}\right)\left(\theta C_x^2 - 4\rho_{yx}C_yC_x\right)\right] \tag{16}$$

Now, after squaring of (15) and neglecting the higher power terms of $e's$, we have

$$(\zeta-\bar{Y})^2 \cong \bar{Y}^2\left[(\phi-1)+\phi\left(e_y+\frac{1}{2}\theta e_x+\frac{1}{2}\theta e_y e_x-\frac{1}{8}\theta^2 e_x^2\right)\right]^2 \tag{17}$$

Taking expectation of (17), we get the MSE of the suggested estimator ζ as

$$MSE(\zeta) = \bar{Y}^2\left[(\phi-1)^2+\phi^2 A+2(\phi^2-\phi)B\right] \tag{18}$$

where $A=\left(\frac{1}{r}-\frac{1}{N}\right)\left(C_y^2+\frac{1}{4}\theta^2C_x^2+\theta\rho_{yx}C_yC_x\right)$, $B=\left(\frac{1}{r}-\frac{1}{N}\right)\left(-\frac{1}{8}\theta^2C_x^2+\frac{1}{2}\theta\rho_{yx}C_yC_x\right)$. Differentiating (18) with respect to ϕ, and its equating to zero respectively, we get the optimum value of ϕ, is given by

$$\phi_{opt} = \frac{1+\left(\frac{1}{r}-\frac{1}{N}\right)\left(-\frac{1}{8}\theta^2C_x^2+\frac{1}{2}\theta\rho_{yx}C_yC_x\right)}{1+\left(\frac{1}{r}-\frac{1}{N}\right)\left(C_y^2+2\theta\rho_{yx}C_yC_x\right)} \tag{19}$$

After substituting the optimum value of ϕ, i. e., ϕ_{opt} in equation (18), we obtain the minimum MSE of the suggested estimator ζ, is given as

$$MSE(\zeta)_{opt} = \left[1-\frac{\left(1+\left(\frac{1}{r}-\frac{1}{N}\right)\left(-\frac{1}{8}\theta^2C_x^2+\frac{1}{2}\theta\rho_{yx}C_yC_x\right)\right)^2}{1+\left(\frac{1}{r}-\frac{1}{N}\right)\left(C_y^2+2\theta\rho_{yx}C_yC_x\right)}\right]\bar{Y}^2 \tag{20}$$

6. Simulation Study

We have considered the four data sets for the sample population between 25% to 50%, response rate between 60% to 94% with different correlation coefficient. The percent relative efficiency of the suggested estimator engaged in simulation study. The PREs of the suggested estimator ζ with respect to the mean imputation method, Singh and Deo (2003) estimator and Adapted estimator (Bahl and Tuteja (1991)) are obtained as

$$PRE_1 = \frac{V(\bar{y}_r)}{MSE(\zeta)_{opt}}\times 100 \tag{21}$$

$$PRE_2 = \frac{MSE(\bar{y}_{SD})}{MSE(\zeta)_{opt}}\times 100 \tag{22}$$

$$PRE_3 = \frac{MSE(\bar{y}_{AE})}{MSE(\zeta)_{opt}}\times 100 \tag{23}$$

Table 1: Description of Data sets

Parameters	Data set 1 Maddala (1977)	Data set 2 Pandey and Dubey (1988)	Data set 3 Singh, S. (2003)	Data set 4 Swain (2013)
N	16	20	30	50
n	4	8	10	15
r	3	(6, 7)	(6,7,8,9)	(10, 12, 14)
$\bar{Y}$	7.6375	19.55	384.2	6.5945
$\bar{X}$	75.4343	18.8	67.267	55
C_y	0.2278	0.3552	0.1559	0.2134
C_x	0.0986	0.3943	0.1371	0.2968
ρ_{yx}	-0.6823	-0.9199	-0.8552	-0.8628

Table 2: Percent relative efficiency of the suggested estimator ζ over the estimators $\bar{y}_r$, $\bar{y}_{SD}$ and $\bar{y}_{AE}$ respectively under the four different Data sets.

$Dataset$	N	n	r	PRE_1	PRE_2	PRE_3
1	16	4	3	133.898	117.282	100.626
2	20	8	6	349.680	248.514	100.332
			7	349.152	294.760	100.180
3	30	10	6	226.579	143.789	99.9824
			7	226.579	161.787	99.9823
			8	226.579	181.421	99.9822
			9	226.579	202.924	99.9821
4	50	15	10	353.940	285.272	100.376
			12	353.641	310.309	100.291
			14	353.429	338.191	100.231

7. Analysis of Simulation Study

From Tables (1 - 2), the following interpretation can be read out:
(1) From Table 1 presents the parameters of the four data sets for different correlation coefficient. We are taking different values for n and r.
(2) From Table 2 it is observed that

(a) For a 25% sample population with response rate is 75%, the PRE of the suggested estimator ζ with respect to the other existing estimators like as the mean imputation method remains 133.898%, Singh and Deo ($\bar{y}_{SD}$) estimator remains 117.282 % and Adapted estimator remains 100.626%.

(b) For a 40% sample population with response rate between 75% to 87%, the PRE of the suggested estimator ζ with respect to the other existing estimators like as the mean imputation method remains 349.152% to 349.680%, Singh and Deo ($\bar{y}_{SD}$) estimator remains 248.514% to 294.760% and Adapted estimator remains 100.180% to 100.332%.

(c) For a 33% sample population with response rate between 60% to 90%, the PRE of the suggested estimator ζ with respect to the other existing estimators like as the mean imputation method remains 226.579% to 226.579%, Singh and Deo ($\bar{y}_{SD}$) estimator remains 143.789% to 202.924% and Adapted estimator remains 99.9821% to 99.9824 %.

(d) For a 30% sample population with response rate between 67% to 94%, the PRE of the suggested estimator ζ with respect to the other existing estimators like as the mean imputation method remains 353.429% to 353.940%, Singh and Deo ($\bar{y}_{SD}$) estimator remains 285.272% to 338.191% and Adapted estimator remains 100.231% to 100.376 %.

8. Conclusions

From the above analysis, it is observed that the suggested estimator is more efficient than the mean imputation method, Singh and Deo (2003) estimator and Adapted estimator (Bahl and Tuteja (1991)). Hence, it can be recommended for future use.

Acknowledgements

The author is grateful to the reviewers for their constructive comments and valuable suggestions regarding improvement of the article.

REFERENCES

BAHL, S., TUTEJA, R. K., (1991). Ratio and Product type exponential estimator. Journal of Information and Optimization Sciences, 12 (1), pp. 159–164.

MADDALA, G. S., (1977). Econometrics. McGraw Hills Pub. Co., New York.

MURTHY, M. N., (1967). Sampling theory and methods. Statistical Publishing Society, Calcutta, India.

PANDEY, B. N., DUBEY, V., (1988). Modified product estimator using coefficient of variation of auxiliary variable. Assam Stat. Review, 2, pp. 64–66.

PRASAD, S., (2016). A study on new methods of ratio exponential type imputation in sample surveys. Hacettepe Journal of Mathematics and Statistics, DOI:10.15672/HJMS.2016.392.

PRASAD, S., (2017). Ratio exponential type imputation in sample surveys. Model Assisted Statistics and Application, 12 (2), pp. 95–106.

SINGH, S., DEO, B., (2003). Imputation by power transformation. Statistical Papers, 44, pp. 555–579.

SINGH, S., (2003). Advanced sampling theory with applications. Klumer Academic Publishers, the Netherlands, Vol. 1.

SWAIN, A. K. P. C., (2013). On some modified ratio and product type estimators-Revisited. Revista Investigacion Operacional, 34 (1), pp. 35–57.