Exponential Ratio Type Estimators of Population Mean Under Non Response

Paramdeep Kaur and Lovleen Kumar Grover

Department of Mathematics, Guru Nanak Dev University, Amritsar-143005, Punjab, India,
Corresponding author: lovleen_2@yahoo.co.in

Abstract

This paper proposes some exponential ratio type estimators of population mean under the situations when certain observations for some sampling units are missing. These missing observations may be for either auxiliary variable or study variable or both of these variables. The biases and mean square errors of these proposed estimators have been derived, up to first order of approximation. These proposed estimators are compared theoretically with that of the existing ratio type estimators defined by Toutenberg and Srivastava (1998). We obtained the conditions for which our proposed estimators are better than the corresponding ratio type estimators of Toutenberg and Srivastava (1998). To verify these theoretical results obtained, a simulation study is carried out finally.

Keywords: auxiliary variable, bias, mean square error, non-response, simple random sampling without replacement, study variable

1. Introduction

In survey sampling situations, auxiliary information is oftenly used to improve the precision or accuracy of the estimator of unknown population parameter of interest under the assumption that all the observations in the sample are available. Then, it is conventional to estimate unknown population mean of study variable using ratio estimator provided that there is a positive correlation between study variable and auxiliary variable. (see Cochran (1940)). But in many survey sampling situations, it is not true, that is, the case of incomplete information which may arises due to some non-response in the given sample. There are various practical reasons for this incomplete information due to non-response like: unwillingness of the respondent to answer some particular questions, accidental loss of information caused by unknown factors, failure on the part of investigator to collect correct information, etc. Such type of incomplete information is very common in the studies related to medical research, market research surveys, opinion polls, socio economic investigations, etc. When information about all the units is not available then traditional complete data estimating procedures could not be used straight forwardly to analyze the data.

We have seen that Toutenberg and Srivastava (1998) considered various ratio type estimators of population mean, under different situations, when some of the observations on either the study variable or auxiliary variable or both of these variables are missing. In this paper, we have assumed the same situations of non response as considered by them. In Section 2, we propose the corresponding exponential ratio type estimators of population mean for these situations. In Section 3, we obtain the expressions for the biases and mean square errors (MSE) of these proposed estimators. In Section 4, the comparisons (with respect to biases and mean square errors) have been made for the proposed estimators with the corresponding existing estimators of Toutenberg and Srivastava (1998). Finally, in Section 5, a simulation study has been performed to support the theoretical results obtained earlier in this paper.

2. Notations and the proposed estimators

Let y and x denote the positive valued study variable and positive valued auxiliary variable respectively. Assume that there is a positive correlation between y and x. Let (Y_i, X_i); $i=1,2,3, \ldots, N$ denote the values of bivariate (y,x) on the i^{th} unit of population of size N. Consider a sample, say 's', of size n is drawn with simple random sampling without replacement (SRSWOR) from this population. Now, the problem of interest is to estimate the unknown population mean $\bar{Y}$ of study variable y when the population mean $\bar{X}$ of auxiliary variable x is assumed to be known.

It is assumed that $(n-p-q)$ observations of (y,x), namely $(y_1,x_1),(y_2,x_2),\ ..,(y_{n-p-q},x_{n-p-q})$ measured on selected units in the sample are completely available. In addition to these available observations, let $x_1^*,x_2^*,......x_p^*$ denote the available observations of x variable on other p units in the sample but the corresponding observations of y variable are missing on these p sample units. Similarly, we have a set of other q available observations of y variable, namely $y_1^{**},y_2^{**},...,y_q^{**}$ in the sample but the associated values of x variable are missing on these q sample units. For $(n-p-q)$ sampling units, the observations $(y_i,x_i)^{'s}$ of (y,x) are known which constitutes a sub sample 's_1'. For the other p sampling units, the observations $x_i^{*\,'s}$ of x are known but the corresponding observations of y are not known which constitutes a sub sample 's_2'. For the remaining q sampling units the observations $y_i^{**'s}$ of y are known but corresponding observations of x are not known which constitutes a sub sample 's_3'

We note that $s=s_1\cup s_2\cup s_3$ and $s_i\cap s_j=\phi(\text{an empty set})$ for $i,j=1,2,3$ and $i\neq j.$

Here, the quantities p and q denote the numbers of sampling units in the sample 's' with incomplete observations on bivariate (y,x) and these must be random in nature.

Let

$$\left.\begin{array}{l}\bar{x}=\dfrac{1}{n-p-q}\sum_{i\in s_1}x_i,\quad \bar{y}=\dfrac{1}{n-p-q}\sum_{i\in s_1}y_i,\qquad \bar{x}^*=\dfrac{1}{p}\sum_{i\in s_2}x_i^*\\ \bar{y}^{**}=\dfrac{1}{q}\sum_{i\in s_3}y_i^{**},\qquad \bar{y}_A=\dfrac{(n-p-q)\bar{y}+q\bar{y}^{**}}{(n-p)},\ \bar{x}_A=\dfrac{(n-p-q)\bar{x}+p\bar{x}^*}{(n-q)}\end{array}\right]\tag{1}$$

Toutenberg and Srivastava (1998) defined the following four ratio type estimators of $\bar{Y}$:

$$\bar{y}_{r_1}=\frac{\bar{y}}{\bar{x}}\bar{X},\quad \bar{y}_{r_2}=\frac{\bar{y}}{\bar{x}_A}\bar{X}=\frac{(n-q)\bar{y}}{(n-p-q)\bar{x}+p\bar{x}^*}\bar{X}$$

$$\bar{y}_{r_3}=\frac{\bar{y}_A}{\bar{x}}\bar{X}=\frac{(n-p-q)\bar{y}+q\bar{y}^{**}}{(n-p)\bar{x}}\bar{X}\ \ \bar{y}_{r_4}=\frac{\bar{y}_A}{\bar{x}_A}\bar{X}=\frac{(n-q)}{(n-p)}\left[\frac{(n-p-q)\bar{y}+q\bar{y}^{**}}{(n-p-q)\bar{x}+p\bar{x}^*}\right]\bar{X}$$

In the ordinary circumstances, when there is no non response then Bahl and Tuteja (1991) introduced an exponential ratio type estimator of $\bar{Y}$ which is better than mean per unit estimator of $\bar{Y}$ even for the low positive correlation between y and x. On the other hand, the ordinary ratio estimator of $\bar{Y}$ (due to Cochran (1940)) is better than mean per unit estimator for high positive correlation between y and x. On taking this advantage of exponential ratio type estimators and then considering the concept of ratio type estimators defined by Toutenberg and Srivastava (1998), we have got a motivation to propose the following exponential ratio type estimators of $\bar{Y}$:

$$\bar{y}_{\mathrm{Re}_1}=\bar{y}\exp\left(\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right),\ \bar{y}_{\mathrm{Re}_2}=\bar{y}\exp\left(\frac{\bar{X}-\bar{x}_A}{\bar{X}+\bar{x}_A}\right)$$

$$\bar{y}_{\mathrm{Re}_3} = \bar{y}_A \exp\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}, \quad \bar{y}_{\mathrm{Re}_4} = \bar{y}_A \exp\frac{\bar{X}-\bar{x}_A}{\bar{X}+\bar{x}_A}$$

3. Biases and Mean Square errors of Proposed Estimators

To find biases and mean square errors of the proposed estimators, we proceed as follows.

$$\text{Let } U = \frac{\bar{x}}{\bar{X}} - 1, \quad V = \frac{\bar{y}}{\bar{Y}} - 1, \quad U^* = \frac{\bar{x}^*}{\bar{X}} - 1, \quad V^{**} = \frac{\bar{y}^{**}}{\bar{Y}} - 1. \tag{2}$$

Now we state the following lemma.

Lemma 3.1: Under SRSWOR, we have the following expectations:

$$\left.\begin{array}{l} E(U) = E(V) = E(U^*) = E(V^{**}) = 0, \; E(U^2) = f_{p+q}C_x^2, \, E(V^2) = f_{p+q}C_y^2, \\ E(UV) = f_{p+q}\rho C_y C_x, \; E(U^{*2}) = f_p C_x^2, \; E(V^{**2}) = f_q C_y^2, \\ E(UU^*) = E(U^*V) = E(VV^{**}) = E(UV^{**}) = 0 \end{array}\right]$$

Where $f_{p+q} = E_1\left(\frac{1}{n-p-q}\right) - \frac{1}{N}, \; f_p = E_1\left(\frac{1}{n-p}\right) - \frac{1}{N}, \; f_q = E_1\left(\frac{1}{n-q}\right) - \frac{1}{N},$

E_1 denotes the unconditional expectation based on the whole sample *s*,

$$C_x^2 = \frac{1}{(N-1)\bar{X}^2}\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2, \quad C_y^2 = \frac{1}{(N-1)\bar{Y}^2}\sum_{i=1}^{N}\left(Y_i - \bar{Y}\right)^2$$

$$\rho = \text{corr }(y,x) = \frac{1}{(N-1)\bar{X}\bar{Y}C_x C_y}\sum_{i=1}^{N}\left(Y_i - \bar{Y}\right)\left(X_i - X\right)$$

Remark 3.1: We note that

$f_p \le f_{p+q}$ and $f_q \le f_{p+q}$ always hold good, where as $f_p \lessgtr f_q$, hold good according as $p \lessgtr q$.

Theorem 3.1: The expressions of biases and mean square errors of the four proposed exponential ratio type estimators of $\bar{Y}$, up to first order of approximation, are

$$Bias\left(\bar{y}_{\mathrm{Re}_1}\right) = \bar{Y} f_{p+q}\left(\frac{3}{8}C_x^2 - \frac{1}{2}\rho C_y C_x\right); \quad MSE\left(\bar{y}_{\mathrm{Re}_1}\right) = \bar{Y}^2 f_{p+q}\left(\frac{C_x^2}{4} + C_y^2 - \rho C_y C_x\right)$$

$$Bias\left(\bar{y}_{\mathrm{Re}_2}\right) = \bar{Y} f_q\left(\frac{3C_x^2}{8} - \frac{1}{2}\rho C_y C_x\right); \quad MSE\left(\bar{y}_{\mathrm{Re}_2}\right) = \bar{Y}^2\left(\frac{1}{4}C_x^2 f_q + C_y^2 f_{p+q} - \rho C_y C_x f_q\right)$$

$$Bias\left(\bar{y}_{\mathrm{Re}_3}\right) = \bar{Y}\left(\frac{3}{8}C_x^2 f_{p+q} - \frac{1}{2}\rho C_y C_x f_p\right); \quad MSE\left(\bar{y}_{\mathrm{Re}_3}\right) = \bar{Y}^2\left(\frac{1}{4}C_x^2 f_{p+q} + C_y^2 f_p - \rho C_y C_x f_p\right)$$

$$Bias\left(\bar{y}_{\mathrm{Re}_4}\right)=\bar{Y}\left(\frac{3}{8}C_x^2 f_q-\frac{1}{2}\rho C_y C_x f_p\right);\;\; MSE\left(\bar{y}_{\mathrm{Re}_4}\right)=\bar{Y}^2\left(\frac{C_x^2}{4} f_q+f_p C_y^2-\rho C_y C_x f_p\right)$$

4. Comparison of Proposed Estimators With Existing Estimators

To compare biases and mean square errors of the proposed estimators with the estimators defined by Totenberg and Srivastava (1998), we require the expressions of their biases and mean square errors, up to first order of approximation.

Table 4.1: Biases and mean square errors of existing estimators, up to first order of approximation

Estimator	$Bias$ (.)	MSE (.)
$\bar{y}_{r_1}$	$\bar{Y}\, f_{p+q}\left(C_x^2-\rho C_x C_y\right)$	$\bar{Y}^2 f_{p+q}\left(C_x^2+C_y^2-2\rho C_y C_x\right)$
$\bar{y}_{r_2}$	$\bar{Y}\left(f_q C_x^2-\rho C_y C_x f_q\right)$	$\bar{Y}^2\left(f_{p+q}C_y^2+f_q C_x^2-2\rho C_y C_x f_q\right)$
$\bar{y}_{r_3}$	$\bar{Y}\left(f_{p+q}C_x^2-\rho C_y C_x f_p\right)$	$\bar{Y}^2\left(C_x^2 f_{p+q}+C_y^2 f_p-2\rho C_y C_x f_p\right)$
$\bar{y}_{r_4}$	$\bar{Y}\left(f_q C_x^2-\rho C_y C_x f_p\right)$	$\bar{Y}^2\left(f_p C_y^2+f_q C_x^2-2\rho C_y C_x f_p\right)$

Remark 4.1: While comparing biases and mean square errors, we have taken p and q as fixed quantities in the given sample, so we must have

$$E_1\left(\frac{1}{n-p}\right)=\frac{1}{n-p},\, E_1\left(\frac{1}{n-q}\right)=\frac{1}{n-q},\, E_1\left(\frac{1}{n-p-q}\right)=\frac{1}{n-p-q}$$

Remark 4.2: Reddy (1978) has shown that the values of parameter $K=\rho\dfrac{C_y}{C_x}$ remain stable in any repetitive survey. So while comparing biases and mean square errors of various estimators, we shall try to find the conditions on the values of K under which one estimator is superior to the other estimator. In the present situations, we also note that the value of K always lies in the interval $0<K<\infty$.

Theorem 4.1: Up to the terms of order n^{-1}, we have

$$\left|Bias\left(\bar{y}_{\mathrm{Re}_1}\right)\right|<\left|Bias\left(\bar{y}_{r_1}\right)\right| \quad \text{if} \quad K\in\left(0,\frac{88}{96}\right)\cup\left(\frac{5}{4},\infty\right) \tag{3}$$

$$\left|Bias\left(\bar{y}_{\mathrm{Re}_2}\right)\right|<\left|Bias\left(\bar{y}_{r_2}\right)\right| \quad \text{if} \quad K\in\left(0,\frac{88}{96}\right)\cup\left(\frac{5}{4},\infty\right) \tag{4}$$

$$\left|Bias\left(\bar{y}_{\mathrm{Re}_3}\right)\right| < \left|Bias\left(\bar{y}_{r_3}\right)\right| \quad \text{if} \quad K \in \left(0, \frac{88 f_{p+q}}{96 f_p}\right) \cup \left(\frac{5 f_{p+q}}{4 f_p}, \infty\right) \tag{5}$$

$$\left|Bias\left(\bar{y}_{\mathrm{Re}_4}\right)\right| < \left|Bias\left(\bar{y}_{r_4}\right)\right| \quad \text{if } K \in \left(0, \frac{88}{96}\frac{f_q}{f_p}\right) \cup \left(\frac{5}{4}\frac{f_q}{f_p}, \infty\right) \tag{6}$$

Theorem 4.2: Up to the term of order n^{-1}, we have

$$MSE\left(\bar{y}_{\mathrm{Re}_1}\right) < MSE\left(\bar{y}_{r_1}\right) \quad \text{if} \quad K < \frac{3}{4} \tag{7}$$

$$MSE\left(\bar{y}_{\mathrm{Re}_2}\right) < MSE\left(\bar{y}_{r_2}\right) \quad \text{if} \quad K < \frac{3}{4} \tag{8}$$

$$MSE\left(\bar{y}_{\mathrm{Re}_3}\right) < MSE\left(\bar{y}_{r_3}\right) \quad \text{if} \quad K < \frac{3}{4}\frac{f_{p+q}}{f_p} \tag{9}$$

$$MSE\left(\bar{y}_{\mathrm{Re}_4}\right) < MSE\left(\bar{y}_{r_4}\right) \quad \text{if} \quad K < \frac{3}{4}\frac{f_q}{f_p} \tag{10}$$

From the above two theorems, we see that our proposed estimators are superior to the existing estimators under some very simple conditions.

5. A Simulation Study

To support the facts proved in earlier sections of this paper, we perform a simulation study here. For this purpose, we have taken an empirical population of size 34 from Singh and Chaudhary (2002) [page No. 177]. In this population, variable $'y'$ is area under wheat in 1973 and variable $'x'$ is area under wheat in 1971. For this population, we have following requisite parameters:

$$\bar{X} = 856.41,\ \bar{Y} = 208.88,\ C_x = 0.86,\ C_y = 0.72,\ \rho = 0.45,\ K = 0.38$$

We have simulated sample (with SRSWOR) 50 times from the above fixed population using R software (version 2.14.0). We have taken the following different values of triplet (n, p, q) for sample size 17:

p=3, q=3	p=3, q=4	p=3, q=5	p=4,q=3	p=5,q=3

In the next table, we have mentioned the variances of values of various estimators, considered in this paper, obtained for the simulated 50 different samples drawn from the given population on taking sample size 17 and various values of missingness rates.

Table 5.1: Variances of various estimators for sample size $'n' = 17$

Missingness rate	**Variances of various estimators for some fixed values of *p* and *q***							
	$\bar{y}_{r_1}$	$\bar{y}_{\mathrm{Re}_1}$	$\bar{y}_{r_2}$	$\bar{y}_{\mathrm{Re}_2}$	$\bar{y}_{r_3}$	$\bar{y}_{\mathrm{Re}_3}$	$\bar{y}_{r_4}$	$\bar{y}_{\mathrm{Re}_4}$
p=3, q=3	1668.53	1227.91	1374.60	1078.75	2154.12	1251.30	1728.38	**1071.26**
p=3,q=4	1763.77	1216.27	1528.68	1128.63	2404.94	1321.03	1801.96	**1097.19**
p=3,q=5	1942.14	1372.84	1693.95	1271.82	2318.14	1224.22	1842.73	**1036.72**
p=4, q=3	1910.87	1324.83	1584.99	**1212.96**	2370.69	1368.23	1980.46	1239.96
p=5,q=3	2004.64	1481.48	1841.21	1476.93	2394.40	**1469.55**	2253.30	1474.10

Discussion, conclusion and Interpretations:

For the given fixed population, $K = 0.38$, which satisfies all the conditions, obtained in the results (7) to (10), for all values of triplet (n, p, q), considered in the above table. Therefore, the variances of all the proposed exponential ratio type estimators are less than that of all the corresponding existing ratio type estimators of Toutenberg and Srivastava (1998). It can be verified through the above table.

6. References

1) Bahl, S. and Tuteja, R.K. (1991). Ratio and product type exponential estimators. *Journal of Information and Optimization Sciences*, 12, 159-163.
2) Cochran W.G. (1940). The estimation of yields of cereal experiments by sampling for the ratio gain to total produce. *Journal of Agriculture Society*, 30, 262-275.
3) Singh, D., Chaudhary, F.S. (2002). Theory and analysis of sample survey designs. *New Age International (P) Limited Publishers, New Delhi, India*.
4) Toutenberg, H. and Srivastava, V.K. (1998). Estimation of ratio of population means in survey sampling when some observations are missing. *Metrika*, 48, 177-187.
5) Reddy, V.M. (1978). A study on the use of prior knowledge on certain population parameters in estimation. *Sankhya, Sr C*, 40, 29-37.