Springer**Plus**

CrossMark

Some classes of estimators in the presence of non-response using auxiliary attribute

Saba Riaz[1†] and Md. Abud Darda[2*†] (ID)

*Correspondence:
dardabd@gmail.com
†Saba Riaz and Md. Abud
Darda contributed equally to
this article
[2] Natural Science Academic
Group, National University,
Gazipur 1704, Bangladesh
Full list of author information
is available at the end of the
article

Abstract

In this paper, possible solutions of problem of non-response in the variable of interest are proposed when information about an auxiliary attribute is available. By taking motivation from the previous work, modified classes have been suggested for estimating population mean. Two new generalized classes of estimators are presented along with their asymptotic biases and variances. Efficacy analysis of the suggested classes is acquired with the usual regression estimator. Two real examples have been provided to show the efficiency of the proposed design approach and comparison of suggested estimators with the linear regression estimator.

Keywords: Proportion, Bias, Mean square error, Ratio function, Exponential function, Efficiency

Introduction and literature review

In sample surveys while estimating an unknown population parameter of the study variable, use of the auxiliary information increases the efficiency of the estimates. In many real scenarios, some auxiliary attributes are highly correlated with the study variable, for instance, person's weight and gender, cow's amount of produced milk and breed, crop production and seed type, income and ownership of a house, etc. Many authors have suggested estimators based on auxiliary attributes in simple random sampling or two phase sampling [see Naik and Gupta (1996), Jhajj et al. (2006), Singh et al. (2008), Shabbir and Gupta (2010), Abd-Elfattah et al. (2010), Singh and Solanki (2012), Koyuncu (2012) and references cited therein]. Sometimes people might refuse to reveal their weight or income, or some information are missing because of respondents' non availability at time of the survey. These situations are related to non-response problem when objects are unavailable or people refuse to answer. In the literature, much work has been done in the existence of non-response problem in the variable of interest and in single or two phase sampling scheme [see Khare and Srivastava (1997), Okafor and Lee (2000), Singh and Kumar (2010), Riaz et al. (2014) and references cited therein].

Springer Open

Diana et al. (2012) suggested regression type estimators using auxiliary attribute in presence of non-response. According to best of our knowledge little attention is given to the problem of non-response using information on auxiliary attributes.

The aim of this paper is to provide solution to this important problem. At this end, two generalized classes are suggested for the unknown population mean along with the supplementary information about the proportion P of the population units having the auxiliary attribute ϕ. Taking motivation from Shabbir and Gupta (2007), Koyuncu (2012) and Singh and Solanki (2012), some modified classes are also suggested in the existence of missing information.

Let a finite population U of N distinctive units. Let Y be the variable of interest having values y_i, $i = (1, \ldots, N)$ with unknown mean $\bar{Y} = \sum_{i=1}^{N} y_i / N$ and unknown variance $S_y^2 = \sum_{i=1}^{N} (y_i - \bar{Y})^2 / (N - 1)$ assuming that non-response occurs in Y. Let ϕ be the auxiliary attribute correlated with Y having values ϕ_i, $i = (1, \ldots, N)$. Consider $\phi_i = 1$, if the ith unit of the population has attribute ϕ and 0, otherwise. Let $A = \sum_{i=1}^{N} \phi_i$ and $a = \sum_{i=1}^{n} \phi_i$ be the total number of units in the population and in the sample having attribute ϕ, $P = (A/N)$ and $\hat{P} = (a/n)$ denote the proportion of units in the population and in the sample having attribute ϕ. Let P is known and used to estimate the mean $\bar{Y}$. Hansen and Hurwitz (1946) suggested the following sub-sampling technique in presence of non-response. Let a large sample u of size n ($n < N$) by simple random sampling without replacement (SRSWOR) to collect information on Y. Assuming at first phase, Y can be observed only for n_1 units out of n and, the remaining $n_2 = n - n_1$ units are taken as non-response. A sub-sample of size $r = n_2/k$, $k > 1$ is selected from non-response units where r would be an integer or must be rounded. Assuming that all r selected units give full response on second call. In this fashion, the population is said to be divided into two groups U_1 and U_2 of sizes N_1 and N_2, where U_1 is a group of respondents that would give response on the first call at second phase and U_2 is non-respondents group which would respond on the second call. Obviously N_1 and N_2 are unknown quantities.

Considering the above situations, Hansen and Hurwitz (1946) have suggested the following estimator for population mean,

$$\bar{y}^* = d_1 \bar{y}_1 + d_2 \bar{y}_{2r}, \tag{1}$$

where

$$\bar{y}_1 = \frac{\sum_{i=1}^{n_1} y_i}{n_1}, \quad \bar{y}_{2r} = \frac{\sum_{i=1}^{r} y_i}{r}, \quad d_1 = \frac{n_1}{n} \quad \text{and} \quad d_2 = \frac{n_2}{n}.$$

As well known $\bar{y}^*$ is an unbiased estimator of $\bar{Y}$

$$E(\bar{y}^*) = D_1 \bar{Y}_1 + D_2 \bar{Y}_2 = \bar{Y},$$

where

$$\bar{Y}_1 = \frac{\sum_{i=1}^{N_1} y_i}{N_1}, \quad \bar{Y}_2 = \frac{\sum_{i=1}^{N_2} y_i}{N_2}, \quad D_1 = \frac{N_1}{N} \quad \text{and} \quad D_2 = \frac{N_2}{N}.$$

The variance of $\bar{y}^*$ is given by

$$\text{Var}(\bar{y}^*) = [\theta S_y^2 + \lambda S_{y(2)}^2] = \widetilde{S}_y^2, \tag{2}$$

where

$$S_{y(2)}^2 = \frac{\sum_{i=1}^{N_2}(y_i - \bar{Y}_2)^2}{N_2 - 1}, \quad \theta = \left(\frac{1}{n} - \frac{1}{N}\right) \quad \text{and} \quad \lambda = \frac{N_2(k-1)}{nN}.$$

We can write (2) as

$$\mathrm{Var}(\bar{y}^*) = \widetilde{S}_y^2 = \bar{Y}^2 \widetilde{C}_y^2.$$

If P is known, we can first of all consider the regression estimator of $\bar{Y}$

$$\bar{y}_{\mathrm{reg}}^* = \bar{y}^* + w(P - \hat{P}), \tag{3}$$

where w is an unknown constant to be selected properly.

The mean square error (MSE) of $\bar{y}_{\mathrm{reg}}^*$ is minimum when

$$w = \frac{S_{y\phi}}{S_\phi^2} = w^o(\text{say}).$$

The minimum MSE of $\bar{y}_{\mathrm{reg}}^*$ is given by

$$\mathrm{minMSE}\left(\bar{y}_{\mathrm{reg}}^*\right) = \theta S_y^2\left(1 - \rho_{pb}^2\right) + \lambda S_{y(2)}^2, \tag{4}$$

where

$$\rho_{pb}^2 = \frac{S_{y\phi}^2}{S_y^2 S_\phi^2}, \quad S_\phi^2 = \frac{\sum_{i=1}^{N}(\phi_i - P)^2}{N - 1} \quad \text{and} \quad S_{y\phi} = \frac{\sum_{i=1}^{N}(y_i - \bar{Y})(\phi_i - P)}{N - 1}.$$

We can write (4) as

$$\mathrm{minMSE}\left(\bar{y}_{\mathrm{reg}}^*\right) = \theta \bar{Y}^2 C_y^2\left(1 - \rho_{pb}^2\right) + \lambda \bar{Y}^2 C_{y(2)}^2, \tag{5}$$

where

$$C_y^2 = \frac{S_y^2}{\bar{Y}^2}, \quad C_{y(2)}^2 = \frac{S_{y(2)}^2}{\bar{Y}^2}, \quad \rho_{pb}^2 = \frac{C_{yp}^2}{C_y^2 C_p^2}, \quad C_p^2 = \frac{S_\phi^2}{P^2} \quad \text{and} \quad C_{yp} = \frac{S_{y\phi}}{\bar{Y}P}.$$

Modified classes

In this section, we have considered Shabbir and Gupta (2007), Koyuncu (2012) and Singh and Solanki (2012) classes because those are more efficient than the regression estimator. Shabbir and Gupta (2007) suggested a class of ratio estimator for the population mean $\bar{Y}$ using known information of the auxiliary attribute. Later Singh and Solanki (2012) and Koyuncu (2012) proposed classes on the same subject using different known population parameters such as (ρ_{pb}, C_p) etc. Therefore, taking motivation from the work of the just quoted authors, we modify their classes in frame work of non-response on Y.

First modified class

Shabbir and Gupta (2007) proposed the class of estimators

$$\bar{y}_{SG} = \bar{y}\left[w_1 + w_2(P - \hat{P})\right]\left(\frac{P}{\hat{P}}\right), \tag{6}$$

where w_1 and w_2 are suitable weights.

If we assume that there is non-response on Y, the class (6) can be modified as

$$\bar{y}_{M1}^* = \bar{y}^*\left[w_1 + w_2(P - \hat{P})\right]\left(\frac{P}{\hat{P}}\right). \tag{7}$$

For an easy computation of the bias (B) and the mean square error (MSE) of $\bar{y}_{M1}^*$ up to first order approximation, it is convenient to express (7) in terms of δ's

$$\bar{y}_{M1}^* = \bar{Y}\left(1 + \delta_y^*\right)\left[w_1 - w_2 P \delta_\phi\right]\left(1 + \delta_\phi\right)^{-1}, \tag{8}$$

where

$$\delta_y^* = \frac{\bar{y}^* - \bar{Y}}{\bar{Y}},$$

$$\delta_\phi = \frac{\hat{P} - P}{P}$$

and then use Taylor expansion of (8). So, remembering that

$$E(\delta_y^*) = E(\delta_\phi) = 0,$$

$$E(\delta_y^{*2}) = \mathbf{C}_y^2, \quad E(\delta_\phi^2) = \theta C_p^2, \quad E(\delta_y^* \delta_\phi) = \theta C_{yp},$$

we have

$$B(\bar{y}_{M1}^*) = \bar{Y}\left[(w_1 - 1) + \theta(w_1 + w_2 P)\left(C_p^2 - C_{yp}\right)\right] \tag{9}$$

and

$$\mathrm{MSE}(\bar{y}_{M1}^*) = \bar{Y}^2\left[(w_1 - 1)^2 + w_1^2 \widetilde{C}_y^2 + \theta(w_1 + w_2 P)\right.$$
$$\left.\left\{C_p^2(3w_1 + w_2 P - 2) + 2C_{yp}(1 - 2w_1)\right\}\right]. \tag{10}$$

The mean square error of $\bar{y}_{M1}^*$ will be minimum

$$w_1 = \frac{\theta C_p^4 - C_p^2\left(1 + 3\theta C_{yp}\right) + 2\theta C_{yp}^2}{\left[\theta C_p^4 - C_p^2(1 + \mathbf{C}_y^2 + 4\theta C_{yp}) + 4\theta C_{yp}^2\right]} = w_1^o(\text{say})$$

and

$$w_2 = \frac{\left(C_{yp} - C_p^2\right)\left(\theta C_p^2 + C_y^2 - 2\theta C_{yp} - 1\right)}{P\left[\theta C_p^4 - C_p^2(1 + \mathbf{C}_y^2 + 4\theta C_{yp}) + 4\theta C_{yp}^2\right]} = w_2^o(\text{say}),$$

$$\mathrm{minMSE}(\bar{y}_{M1}^*) = \frac{\bar{Y}^2\left[\theta \mathbf{C}_y^2 C_p^4 - \mathbf{C}_y^2 C_p^2(1 + 2\theta C_{yp}) + \theta C_{yp}^2(1 + \mathbf{C}_y^2)\right]}{\left[\theta C_p^4 - C_p^2(1 + \mathbf{C}_y^2 + 4\theta C_{yp}) + 4\theta C_{yp}^2\right]}. \tag{11}$$

Second modified class

Koyuncu (2012) suggested a regression-cum-ratio class

$$\bar{y}_{\mathrm{K}} = \left[w_1 \bar{y} + w_2 (P - \hat{P}) \right] \left(\frac{\eta P + \psi}{\eta \hat{P} + \psi} \right), \tag{12}$$

where w_1 and w_2 having the same expressions defined earlier, η and ψ are either real numbers or functions of the known parameter associated with an auxiliary attribute such as C_p, $\beta_2(\phi)$ and ρ_{pb}. Of course the aim was to increase the performance using more information.

We can modify the class (12), assuming non-response on Y

$$\bar{y}_{\mathrm{M2}}^* = \left[w_1 \bar{y}^* + w_2 (P - \hat{P}) \right] \left(\frac{\eta P + \psi}{\eta \hat{P} + \psi} \right). \tag{13}$$

The bias and the MSE of $\bar{y}_{\mathrm{M2}}^*$ to the first order of approximation can be written as

$$\mathrm{B}(\bar{y}_{\mathrm{M2}}^*) = \bar{Y} \left[(w_1 - 1) + \theta \left(w_1 \tau^2 + w_2 \tau P \right) C_p^2 - \theta w_1 \tau C_{yp} \right] \tag{14}$$

and

$$\begin{aligned}
\mathrm{MSE}(\bar{y}_{\mathrm{M2}}^*) = {} & \bar{Y}^2 (w_1 - 1)^2 + w_1^2 \bar{Y}^2 \mathbf{C}_y^2 + -2\theta w_1 \bar{Y} C_{yp} \left[w_2 P + \bar{Y} P \tau (2w_1 - 1) \right] \\
& + \theta C_p^2 \left[w_1 \bar{Y}^2 \tau^2 (3w_1 - 2) + w_2^2 P^2 + 2(2w_1 - 1) w_2 \bar{Y} P \tau \right],
\end{aligned} \tag{15}$$

where $\tau = \dfrac{\eta P}{\eta P + \psi}$.

The mean square error of $\bar{y}_{\mathrm{M2}}^*$ is minimized for

$$w_1 = \frac{C_p^2 \left(\theta \tau^2 C_p^2 - 1 \right)}{\theta \tau^2 C_p^4 - C_p^2 (1 + \mathbf{C}_y^2) + \theta C_{yp}^2} = w_1^o (\mathrm{say})$$

and

$$w_2 = \frac{\bar{Y} \left[\tau C_p^2 (1 + \theta \tau C_{yp} - \mathbf{C}_y^2) - \theta \tau^3 C_p^4 + C_{yp} (\theta \tau C_{yp} - 1) \right]}{P \left[\theta \tau^2 C_p^4 - C_p^2 (1 + \mathbf{C}_y^2) + \theta C_{yp}^2 \right]} = w_2^o (\mathrm{say}).$$

To emphasize the comparison with the regression estimator (3), we can express the minimum MSE of $\bar{y}_{\mathrm{M2}}^*$ as

$$\mathrm{minMSE}(\bar{y}_{\mathrm{M2}}^*) = \frac{\left(1 - \theta \tau^2 C_p^2 \right) \mathrm{MSE}(\bar{y}_{\mathrm{reg}}^*)}{\left(1 - \theta \tau^2 C_p^2 \right) + \dfrac{\mathrm{MSE}(\bar{y}_{\mathrm{reg}}^*)}{\bar{Y}^2}}. \tag{16}$$

Third modified class

Singh and Solanki (2012) proposed the class

$$\bar{y}_{\mathrm{SS}} = \bar{y} \left[w_1 + w_2 (P - \hat{P}) \right] \left(\frac{\eta P + \delta \psi}{\eta \hat{P} + \delta \psi} \right)^\gamma, \tag{17}$$

where γ is a real number, δ is an integer which takes values $+1$ and -1 for designing the estimators and keeping (w_1, w_2, η, ψ) same as defined before. Note that Shabbir and Gupta (2007) class is a member of this class.

We can modify the class (17) considering incomplete information on Y

$$\bar{y}_{M3}^* = \bar{y}^* \left[w_1 + w_2(P - \hat{P}) \right] \left(\frac{\eta P + \delta \psi}{\eta \hat{P} + \delta \psi} \right)^{\gamma}. \tag{18}$$

The bias and MSE of $\bar{y}_{M3}^*$ to the first order of approximation are given by

$$B(\bar{y}_{M3}^*) = \bar{Y} \left[(w_1 - 1) + \theta C_p^2 \left\{ \gamma v \left(w_2 P + \frac{w_1(\gamma + 1)}{2} \right) - (w_2 P + w_1 \gamma v) k_p \right\} \right] \tag{19}$$

and

$$\text{MSE}(\bar{y}_{M3}^*) = \bar{Y}^2 \left[1 + w_1^2 A^* + w_2^2 B^* + 2w_1 w_2 C^* - 2w_1 D^* - 2w_2 E^* \right], \tag{20}$$

where

$$k_p = \rho_{pb} \frac{C_y}{C_p}, \quad v = \frac{\eta P}{\eta P + \delta \psi},$$
$$A^* = 1 + C_y^2 + \theta \gamma v C_p^2 \left[(2\gamma + 1)v - 4k_p \right],$$
$$B^* = \theta P^2 C_p^2, \quad C^* = 2\theta P C_p^2 (\gamma v - k_p),$$
$$D^* = 1 + \theta \gamma v C_p^2 \left(\frac{(\gamma + 1)v}{2} - k_p \right), \quad E^* = \theta P C_p^2 (\gamma v - k_p).$$

Minimizing the $\text{MSE}(\bar{y}_{M3}^*)$ to achieve optimum values of constants w_1 and w_2

$$w_1 = \frac{B^* D^* - C^* E^*}{A^* B^* - C^{*2}} = w_1^o (\text{say})$$

and

$$w_2 = \frac{A^* E^* - C^* D^*}{A^* B^* - C^{*2}} = w_2^o (\text{say}),$$
$$\text{minMSE}(\bar{y}_{M3}^*) = \bar{Y}^2 \left(1 - \frac{B^* D^{*2} - 2C^* D^* E^* + A^* E^{*2}}{A^* B^* - C^{*2}} \right). \tag{21}$$

Suggested classes

In this section, we have introduced two general classes of estimators for the population mean assuming (as in the previous Section) non-response occurs in the study variable with known information on the auxiliary attribute. The first suggested class is obtained starting from a generalization of the third modified class, defined in (18), while the second class is the result of motivation from Diana et al. (2011).

First class

Let

$$\bar{y}_{S1} = \bar{y}^* (w_1 + w_2 u) g(u), \tag{22}$$

where $u = (P - \hat{P})$, w_1 and w_2 are constants to be chosen properly and g is a generic function that satisfy the following mild conditions

- g is continuous and bounded in a neighborhood of zeros.
- g does not depend on n, N and $(\phi_1, \ldots, \phi_N)$.
- g is a three times differentiable function with continuous and bounded derivatives.

Expanding $g(u)$ using Taylor's series up to order $o_p(u^2)$, the resulting expression for the class is given by

$$\bar{y}_{S1} \cong \bar{y}^*(w_1 + w_2 u)\left[g(0) + g'(0)u + \frac{1}{2}g''(0)u^2\right], \tag{23}$$

where $g(0)$ is a constant term, $g'(0)$ is first order partial derivative of $g(u)$ in zero and $g''(0)$ is second order partial derivative in zero. For sake of simplicity, we can write $g(0) = a_0$, $g'(0) = b_0$ and $\frac{1}{2}g''(0) = c_0$.

Now expressing (23) in terms of δ's

$$\bar{y}_{S1} \cong \bar{Y}\left(1 + \delta_y^*\right)\left(w_1 - w_2 P\delta_\phi\right)\left(a_0 - b_0 P\delta_\phi + c_0 P^2 \delta_\phi^2\right). \tag{24}$$

The bias and the MSE of $\bar{y}_{S1}$ up to the first order of approximation are

$$B(\bar{y}_{S1}) = \bar{Y}\left[(w_1 a_0 - 1) + \theta\left\{(w_1 c_0 + w_2 b_0)P^2 C_p^2 - (w_1 b_0 + w_2 a_0)PC_{yp}\right\}\right] \tag{25}$$

and

$$\begin{aligned}
\text{MSE}(\bar{y}_{S1}) = \bar{Y}^2\Bigg[(w_1 a_0 - 1)^2 + w_1^2 a_0^2 \mathbf{C}_y^2 + \theta\Bigg\{w_1^2\left(b_0^2 P^2 C_p^2 - 2a_0 b_0 PC_{yp}\right) \\
+ (2w_1 b_0 + w_2 a_0)w_2 a_0 P^2 C_p^2 - 2w_2(w_1 a_0 - 1)\left(a_0 PC_{yp} - b_0 P^2 C_p^2\right) \\
+ 2w_1(w_1 a_0 - 1)\left(c_0 P^2 C_p^2 - b_0 PC_{yp}\right)\Bigg\}\Bigg].
\end{aligned} \tag{26}$$

Minimizing (26) with respect to w_1 and w_2, we obtain

$$w_1 = \frac{U_1^* + U_2^*}{a_0\left(T_1^* + T_2^*\right)} = w_1^o(\text{say})$$

and

$$w_2 = \frac{\bar{Y}V_1^* + V_2^*}{a_0^2 P\left(T_1^* + T_2^*\right)} = w_2^o(\text{say}),$$

where

$$T_1^* = a_0^2\left[\bar{Y}^2 C_p^2 + \bar{Y}^2\left(\mathbf{C}_y^2 C_p^2 - \theta C_{yp}^2\right)\right] + \theta \bar{Y}^2 P^2 C_p^4\left(2a_0c_0 - 3b_0^2\right),$$

$$U_1^* = C_p^2\left[a_0^2 + \theta\left(a_0c_0P^2C_p^2 - 2b_0^2P^2C_p^2\right)\right],$$

$$V_1^* = a_0^3 C_{yp} + a_0^2 P\left[b_0\left(\mathbf{C}_y^2 C_p^2 - \theta C_{yp}^2 - C_p^2\right) + \theta c_0 P C_p^2 C_{yp}\right]$$
$$\quad + \theta b_0^3 P^3 C_p^4 - 2\theta a_0 b_0^2 P^2 C_p^2 C_{yp},$$

$$T_2^* = 4\theta a_0 b_0 \bar{Y}^2 P C_p^2 C_{yp} - 3\theta a_0^2 \bar{Y}^2 C_{yp}^2,$$

$$U_2^* = 3\theta a_0 b_0 P C_p^2 C_{yp} - 2\theta a_0^2 C_{yp}^2$$

and

$$V_2^* = 3\theta a_0^2 b_0 P C_{yp}^2 - \theta a_0^3 C_y^2 C_{yp} + \theta a_0 b_0^2 P^2 C_p^2 C_{yp}.$$

Hence, with the help of above notations, one can write minimum MSE of $\bar{y}_{S1}$ as follows

$$\text{minMSE}(\bar{y}_{S1}) = \frac{\bar{Y}^2\left(W_1^* + W_2^*\right)}{a_0^2\left(T_1^* + T_2^*\right)}, \tag{27}$$

where

$$W_1^* = \bar{Y}^2\left[a_0^4\left(\mathbf{C}_y^2 C_p^2 - \theta C_{yp}^2\right) + \theta^2 b_0^2 P^4 C_p^6\left(2a_0c_0 - b_0^2\right)\right.$$
$$\left. -\theta a_0^2 P^2 C_p^2\left\{b_0^2\left(\mathbf{C}_y^2 C_p^2 - \theta C_{yp}^2\right) + \theta c_0^2 P^2 C_p^4\right\}\right]$$

and

$$W_2^* = \bar{Y}^2\left[2\theta^2 a_0 b_0^3 P^3 C_p^4 C_{yp} - 2\theta^2 a_0^2 b_0 c_0 P^3 C_p^4 C_{yp} - \theta a_0^4 \mathbf{C}_y^2 C_{yp}^2\right.$$
$$\left. -3\theta^2 a_0^2 b_0^2 P^2 C_p^2 C_{yp}^2 + 2\theta a_0^3 b_0 P \mathbf{C}_y^2 C_p^2 C_{yp} + 2\theta^2 a_0^3 c_0 P^2 C_p^2 C_{yp}^2\right].$$

Second class

Motivated by Diana et al. (2011), we consider the class

$$\bar{y}_{S2} = \left(w_1 \bar{y}^* + w_2 u\right)g(u), \tag{28}$$

where u and $g(u)$ are explained in earlier section.

If we expand $g(u)$ again using Taylor's series, $\bar{y}_{S2}$ becomes

$$\bar{y}_{S2} \cong \left(\bar{y}^* w_1 + w_2 u\right)\left[g(0) + g'(0)u + \frac{1}{2}g''(0)u^2\right]. \tag{29}$$

The bias and the MSE to the first order of approximation can be written as

$$\text{B}(\bar{y}_{S2}) = \bar{Y}(a_0 w_1 - 1) + \theta\left[(w_1 \bar{Y} c_0 + w_2 b_0)P^2 C_p^2 - w_1 \bar{Y} b_0 P C_{yp}\right] \tag{30}$$

and

$$
\begin{aligned}
\mathrm{MSE}(\bar{y}_{S2}) = {} & \bar{Y}^2(a_0 w_1 - 1)^2 + \Big[w_1^2 a_0^2 \bar{Y}^2 \mathbf{C}_y^2 + \theta P^2 C_p^2 \big\{ w_2^2 a_0^2 \\
& + 2 w_1 a_0 \bar{Y}(2 w_2 b_0 + w_1 c_0 \bar{Y}) + \bar{Y}\big(w_1^2 b_0^2 \bar{Y} - 2 w_2 b_0 - 2 w_1 c_0 \bar{Y}\big) \big\} \\
& - 2\theta w_1 \bar{Y} P C_{yp} \big(w_2 a_0^2 + 2 w_1 a_0 b_0 \bar{Y} - b_0 \bar{Y} \big) \Big].
\end{aligned}
\tag{31}
$$

By minimizing $\mathrm{MSE}(\bar{y}_{S2})$, one can get the optimum values of the constants w_1 and w_2

$$
w_1 = \frac{U_1^*}{a_0 T_1^*} = w_1^o (\text{say})
$$

and

$$
w_2 = \frac{\bar{Y} V_1^*}{a_0^2 P T_1^*} = w_2^o (\text{say}),
$$

Therefore,

$$
\mathrm{minMSE}(\bar{y}_{S2}) = \frac{\bar{Y}^2 W_1^*}{a_o^2 T_1^*}.
\tag{32}
$$

It is observed that Rao (1991) class is a member of Diana et al. (2011). Also in our case, if considered $g(u)$ appears as an identity function, the class (28) reduces to

$$
\bar{y}_{S2(R)}^* = w_1 \bar{y}^* + w_2 u,
\tag{33}
$$

that is the corresponding version of Rao (1991) class when non-response is present.

The bias and the MSE of $\bar{y}_{S2(R)}^*$ can be written as

$$
\mathrm{B}\left(\bar{y}_{S2(R)}^* \right) = \bar{Y}(w_1 - 1)
\tag{34}
$$

and

$$
\mathrm{MSE}\left(\bar{y}_{S2(R)}^* \right) = \bar{Y}^2 (w_1 - 1)^2 + w_1^2 \bar{Y}^2 \mathbf{C}_y^2 + \theta \left(w_2^2 P^2 C_p^2 - 2 w_1 w_2 \bar{Y} P C_{yp} \right).
\tag{35}
$$

The optimum values of constants w_1 and w_2 are

$$
w_1 = \frac{C_p^2}{C_p^2 + \left(\mathbf{C}_y^2 C_p^2 - \theta C_{yp}^2 \right)} = w_1^o (\text{say})
$$

and

$$
w_2 = \frac{\bar{Y} C_{yp}}{P \left[C_p^2 + \left(\mathbf{C}_y^2 C_p^2 - \theta C_{yp}^2 \right) \right]} = w_2^o (\text{say}).
$$

The minimum MSE of $\bar{y}_{S2(R)}^*$ is given by

$$\text{minMSE}\left(\bar{y}_{S2(R)}^*\right) = \frac{\text{MSE}\left(\bar{y}_{\text{reg}}^*\right)}{1 + \dfrac{\text{MSE}\left(\bar{y}_{\text{reg}}^*\right)}{\bar{Y}^2}}. \tag{36}$$

From (36) is clear that the estimator $\bar{y}_{S2(R)}^*$ performs always better than $\bar{y}_{\text{reg}}^*$, in their optimal case.

From (16) and (36), it is easy to see that the minimum MSE of $\bar{y}_{M2}^*$ becomes equal to the minimum MSE of $\bar{y}_{S2(R)}^*$ when $\tau = 0$. One can also observe that regression-cum-ratio estimator (13) may perform better than regression estimator for different choices of τ. Note that the MSE of $\bar{y}_{S2}$ looks like the MSE of Diana et al. (2011), however, their work is related to complete information for the study variable Y and the auxiliary variable X. But the class $\bar{y}_{S2}$ highlights the non-response problem especially when someone estimates the population mean using information of the auxiliary attribute.

Choice of function *g*

The performance of the proposed classes $\bar{y}_{S1}$ and $\bar{y}_{S2}$ depends upon the selection of function *g*. Careful choice for *g* is a crucial factor and it requires deep insight both from theoretical and practical point of view. There are many possible choices for *g* but we consider only ratio and exponential function, because they are found good choices from both theoretical and practical point of view.

(i) Consider a ratio type function suggested by Singh and Solanki (2012) assuming $\gamma = 1$

Expanding $g(u)$ by Taylor's theorem, we get

$$g(u) = \left(\frac{\eta P + \delta \psi}{\eta P + \delta \psi - \eta u}\right). \tag{37}$$

$$a_0 = g(0) = 1, \qquad b_0 = g'(0) = \frac{\eta}{\eta P + \delta \psi}, \qquad c_0 = \frac{1}{2}g''(0) = \frac{\eta^2}{(\eta P + \delta \psi)^2}.$$

When $\delta = 1$, $g(u)$ is similar to function considered by Koyuncu (2012).

When $\eta = 1$ and $\psi = 0$ then $g(u)$ become similar to the ratio function suggested by Shabbir and Gupta (2007).

The suggested classes $\bar{y}_{S1}$ and $\bar{y}_{S2}$ become

$$\bar{y}_{S1(1)}^* = \bar{y}^*(w_1 + w_2 u)\left(\frac{\eta P + \delta \psi}{\eta P + \delta \psi - \eta u}\right)$$

and

$$\bar{y}_{S2(1)}^* = (w_1 \bar{y}^* + w_2 u)\left(\frac{\eta P + \delta \psi}{\eta P + \delta \psi - \eta u}\right).$$

If we consider $\eta = 1$ and $\psi = 0$, $\bar{y}_{S1(1)}^*$ is equivalent to $\bar{y}_{M1}^*$ and $\bar{y}_{S2(1)}^*$ is equal to $\bar{y}_{M2}^*$, for $\delta = 1$.

Hence, we can conclude that $(\bar{y}^*_{\text{M1}}, \bar{y}^*_{\text{M3}})$ belong to the class $\bar{y}_{\text{S1}}$ and $\bar{y}^*_{\text{M2}}$ is a member of the class $\bar{y}_{\text{S2}}$.

(ii)   Consider an exponential function

Using Taylor's theorem to expand $g(u)$, we have

$$g(u) = \exp\left(\frac{u}{2P - u}\right). \tag{38}$$

$$a_0 = 1, \quad b_0 = \frac{1}{2P}, \quad c_0 = \frac{3}{8P^2}.$$

Then $\bar{y}_{\text{S1}}$ and $\bar{y}_{\text{S2}}$ can be written as

$$\bar{y}^*_{\text{S1}(2)} = \bar{y}^*(w_1 + w_2 u)\exp\left(\frac{u}{2P - u}\right)$$

and

$$\bar{y}^*_{\text{S2}(2)} = (w_1 \bar{y}^* + w_2 u)\exp\left(\frac{u}{2P - u}\right).$$

The minimum MSE of $\bar{y}^*_{\text{S1}(2)}$ and $\bar{y}^*_{\text{S2}(2)}$ are given by

$$\text{minMSE}\left(\bar{y}^*_{\text{S1}(2)}\right)$$
$$= \frac{\bar{Y}^2\left[\theta^2 C_p^6 + 8\theta C_p^4\left(\theta C_{yp} + 2\mathbf{C}_y^2\right) - 16 C_p^2\left\{4\mathbf{C}_y^2(1 + \theta C_{yp}) + \theta^2 C_{yp}^2\right\} + 64\theta C_{yp}^2\left(1 + \mathbf{C}_y^2\right)\right]}{64\left[4\theta C_{yp}^2 - C_p^2\left(1 + \mathbf{C}_y^2 + 2\theta C_{yp}\right)\right]}$$

and

$$\text{minMSE}\left(\bar{y}^*_{\text{S2}(2)}\right) = \frac{\bar{Y}^2\left[\left(\mathbf{C}_y^2 C_p^2 - \theta C_{yp}^2\right)\left(64\theta - 16\theta C_p^2\right) - \theta C_p^6\right]}{64\left[C_p^2 + \mathbf{C}_y^2 C_p^2 - \theta C_{yp}^2\right]}.$$

Efficiency comparisons

In this section, efficiency of the proposed estimators on the basis of their minimum mean square error has been evaluated by analyzing the performance of estimators, when possible, specially numerically. It is well known that the regression estimator $\bar{y}^*_{\text{reg}}$ is always more efficient than the Hansen and Hurwitz (1946) estimator (for instance, see (2) and (4)). For this reason we make efficiency comparison of the proposed classes with the regression estimator.

From the comparison of (4) with (32), after some computation, one can get

$$\text{MSE}\left(\bar{y}^*_{\text{reg}}\right) - \text{minMSE}(\bar{y}_{\text{S2}}) \geq 0,$$

when

$$\frac{\left[a_0^4 \text{MSE}\left(\bar{y}^*_{\text{reg}}\right) - \theta \bar{Y}^2 P^2 C_p^2\left(b_0^2 - a_0 c_0\right)\right]^2}{a_0^2\left[a_0^2\left\{\bar{Y}^2 + \text{MSE}\left(\bar{y}^*_{\text{reg}}\right)\right\} + \theta \bar{Y}^2 P^2 C_p^2\left(2a_0 c_0 - 3b_0^2\right)\right]} \geq 0.$$

This expression will be certainly > 0 if $\left(2a_0 c_0 - 3b_0^2\right) \geq 0$ and hence $\bar{y}_{S2}$ is more efficient than the regression estimator.

Now making comparison of $\bar{y}_{M2}^*$ and $\bar{y}_{S2(R)}^*$ with $\bar{y}_{reg}^*$

$$\text{MSE}\left(\bar{y}_{reg}^*\right) - \text{minMSE}\left(\bar{y}_{M2}^*\right) \geq 0$$

and

$$\text{MSE}\left(\bar{y}_{reg}^*\right) - \text{minMSE}\left(\bar{y}_{S2(R)}^*\right) \geq 0.$$

Remark It can be observed from (4) and (16) and, from (4) and (36) respectively, that these expressions are always positive. Furthermore, it is not easy to make analytical comparison for $\left(\bar{y}_{M1}^*, \bar{y}_{M3}^*, \bar{y}_{S1}\right)$.

So in this Section, we make numerical comparison of modified and suggested classes using two population data sets as earlier considered by Shabbir and Gupta (2007), Abd-Elfattah et al. (2010) and Koyuncu (2012).

Population I [Source: Sukhatme and Sukhatme (1970), p. 256]

$y =$ Number of villages in the circles.

$\phi =$ A circle consisting of more than five villages.

$N = 89, \qquad n = 23, \qquad \bar{Y} = 3.36, \qquad P = 0.124,$

$C_y = 0.601, \quad C_p = 2.678, \quad \rho_{pb} = 0.766, \quad \beta_2(\phi) = 6.612.$

The non-response rate in the population is considered to be 25 percent, taken as last 22 units of the population

$N_2 = 22, \qquad \bar{Y}_2 = 3.27, \qquad C_{y(2)} = 0.668.$

Population II [Source: Sukhatme and Sukhatme (1970), p. 256]

$y =$ Area (in acres) under the wheat crop within the circles.

$\phi =$ A circle consisting of more than five villages.

$N = 89, \qquad n = 23, \qquad \bar{Y} = 1102, \qquad P = 0.124,$

$C_y = 0.65, \quad C_p = 2.678, \quad \rho_{pb} = 0.624, \quad \beta_2(\phi) = 6.612.$

The non-response units of the population are taken as last 22 units (25 % of N)

$N_2 = 22, \qquad \bar{Y}_2 = 1242.68, \qquad C_{y(2)} = 0.516.$

The comparison is performed in terms of Percent Relative Efficiency (PRE)

$$\text{PRE}\left(\bar{y}_{(\bullet)}^*\right) = \frac{\text{MSE}\left(\bar{y}_{(reg)}^*\right)}{\text{minMSE}\left(\bar{y}_{(\bullet)}^*\right)} \times 100$$

where $\bar{y}_{(\bullet)}^* = \left(\bar{y}_{M1}^*, \bar{y}_{M2}^*, \bar{y}_{M3}^*, \bar{y}_{S1(1)}^*, \bar{y}_{S1(2)}^*, \bar{y}_{S2(R)}^*, \bar{y}_{S2(1)}^*, \bar{y}_{S2(2)}^*\right)$.

From Tables 1 and 2, it is observed that the estimator $\bar{y}_{M2}^*$ with different values of η and ψ performs similar like $\bar{y}_{S2(R)}^*$. It should be noted that the estimator $\bar{y}_{S2(2)}^*$ with exponential function perform better than the estimators $\bar{y}_{M2}^*$ and $\bar{y}_{S2(1)}^*$ with ratio function. After careful

Table 1 PRE of the estimators with respect to $\bar{y}^*_{reg}$ for different values of k for Pop I

Estimator	γ	δ	η	ψ	k			
					2	3	4	5
$\bar{y}^*_{M2}$	–	–	C_p	$\beta_2(\varphi)$	100.94	101.39	101.85	102.30
	–	–	$\beta_2(\varphi)$	C_p	100.95	101.41	101.87	102.33
	–	–	1	C_p	100.94	101.39	101.85	102.30
	–	–	1	$\beta_2(\varphi)$	100.94	101.39	101.85	102.30
$\bar{y}^*_{S2(R)}$	–	–	1	0	100.94	101.39	101.85	102.30
$\bar{y}^*_{S2(1)}$	–	–	1	0	101.22	101.81	102.40	102.99
$\bar{y}^*_{S2(2)}$	–	–	1	0	118.34	114.94	113.55	112.93
$\bar{y}^*_{M3}$	1	1	n	$1 - n/N$	114.65	112.36	111.49	111.18
	1	1	N	P	125.51	120.09	117.82	116.73
	1	1	N	k_p	125.21	119.88	117.65	116.58
$\bar{y}^*_{M1}$	–	–	1	0	126.32	120.65	118.27	117.11
$\bar{y}^*_{S1(1)}$	–	–	1	0	126.32	120.65	118.27	117.11
$\bar{y}^*_{S1(2)}$	–	–	1	0	*135.75*	*126.43*	*122.47*	*120.40*

Table 2 PRE of the estimators with respect to $\bar{y}^*_{reg}$ for different values of k for Pop II

Estimator	γ	δ	η	ψ	k			
					2	3	4	5
$\bar{y}^*_{M2}$	–	–	C_p	$\beta_2(\varphi)$	101.20	101.56	101.93	102.29
	–	–	$\beta_2(\varphi)$	C_p	101.21	101.58	101.95	102.32
	–	–	1	C_p	101.20	101.56	101.93	102.29
	–	–	1	$\beta_2(\varphi)$	101.20	101.56	101.93	102.29
$\bar{y}^*_{S2(R)}$	–	–	1	0	101.20	101.56	101.93	102.29
$\bar{y}^*_{S2(1)}$	–	–	1	0	101.56	102.03	102.51	102.98
$\bar{y}^*_{S2(2)}$	–	–	1	0	115.99	114.28	113.40	112.94
$\bar{y}^*_{M3}$	1	1	n	$1 - n/N$	111.48	110.64	110.29	110.18
	1	1	N	P	118.59	116.52	115.45	114.88
	1	1	N	k_p	118.48	116.43	115.37	114.81
$\bar{y}^*_{M1}$	–	–	1	0	119.10	116.93	115.81	115.20
$\bar{y}^*_{S1(1)}$	–	–	1	0	119.10	116.93	115.81	115.20
$\bar{y}^*_{S1(2)}$	–	–	1	0	*128.32*	*123.85*	*121.40*	*119.92*

analysis of performance of $\bar{y}^*_{M3}$, it is observed that different possible values of η and ψ increase the efficiency of the estimator. For $(\eta = 1,\ \psi = 0)$, the estimators $\bar{y}^*_{M1}$ and $\bar{y}^*_{S1(1)}$ perform similar as expected. The PRE of $\bar{y}^*_{S1(2)}$ is higher than those of $\bar{y}^*_{M1}$, $\bar{y}^*_{M3}$ and $\bar{y}^*_{S1(1)}$ which leads to the conclusion that exponential function may be a better choice than ratio. It can be seen that as the inverse sampling rate k increases, the PREs of the estimators $\left(\bar{y}^*_{M2},\ \bar{y}^*_{S2(R)},\ \bar{y}^*_{S2(1)}\right)$ also increase but the PREs of $\left(\bar{y}^*_{M1},\ \bar{y}^*_{M3},\ \bar{y}^*_{S2(2)},\ \bar{y}^*_{S1(1)},\ \bar{y}^*_{S1(2)}\right)$ decrease.

It has been shown in Singh and Solanki (2012) that the estimator $\bar{y}_{SG}$ with $(\eta = 1,\ \psi = 0)$ performs better than the estimator $\bar{y}_{SS}$ with complete information on Y. The same behavior is observed for $\left(\bar{y}^*_{M1},\ \bar{y}^*_{M3}\right)$ in case of incomplete information on Y, in Table 1 and Table 2. Hence, from practical point of view, $\bar{y}^*_{M1}$ is preferable than $\left(\bar{y}^*_{M2},\ \bar{y}^*_{M3}\right)$ because it is showing higher efficiency by using less auxiliary information as compared with others.

Conclusions

In this paper, two new generalized classes of biased estimators for the population mean have been proposed when information on the auxiliary attribute is available, along with considering the problem of non-response on the study variable. Further, three modified classes of estimators motivated by Shabbir and Gupta (2007), Koyuncu (2012) and Singh and Solanki (2012) have also been considered in presence of non-response. Henceforth, linear regression estimator is considered as benchmark for comparing efficiency of the proposed classes. Our suggested classes $\bar{y}_{S1}$ and $\bar{y}_{S2}$ depend on the choice of function g and for this we consider ratio and exponential functions. Numerical results are reported in Tables 1 and 2 to show superiority of the suggested classes with the regression estimator. The main purpose of this paper is to highlight the non-response problem in the study variable when information of auxiliary attribute is available for estimating the unknown population mean.

Author details
[1] Department of Mathematics and Statistics, Riphah International University, Sector I–14, Islamabad 44000, Pakistan.
[2] Natural Science Academic Group, National University, Gazipur 1704, Bangladesh.

References

Abd-Elfattah AM, El-Sherpieny EA, Mohamed SM, Abdou OF (2010) Improvement in estimating the population mean in simple random sampling using information on auxiliary attribute. Appl Math Comput 215:4198–4202

Diana G, Giordan M, Perri PF (2011) An improved class of estimators for the population mean. Stat Methods Appl 20:123–140

Diana G, Riaz S, Shabbir J (2012) A general class of regression type estimators when auxiliary variable is an attribute. In: Proceedings of the 12th Islamic countries conference on statistical sciences: 19–22 Dec 2012; Qatar, vol 23, pp 277–284. ISOSS: Pakistan

Hansen MH, Hurwitz WN (1946) The problems of non-response in sample surveys. J Am Stat Assoc 41:517–529

Jhajj HS, Sharma MK, Grover LK (2006) A family of estimators of population mean using information on auxiliary attribute. Pak J Stat 22:43–50

Khare BB, Srivastava S (1997) Transformed ratio type estimtors for the population mean in the presence of non-response. Commun Stat Theory Methods 26:1779–1791

Koyuncu N (2012) Efficient estimators of population mean using auxiliary attributes. Appl Math Comput 218(22):10900–10905

Naik VD, Gupta PC (1996) A note on estimating of mean with known population of an auxiliary character. J Indian Soc Agric Stat 48:151–158

Okafor FC, Lee H (2000) Double sampling for ratio and regression estimation with sub-sampling the non-respondents. Surv Methodol 26(2):183–188

Rao TJ (1991) On certain methods of improving ratio and regression estimators. Commun Stat Theory Methods 20:3325–3340

Riaz S, Diana G, Shabbir J (2014) Improved classes of estimators for population mean in presence of non-response. Pak J Stat 30(1):83–100

Shabbir J, Gupta S (2007) On estimating the finite population mean with known population proportion of an auxiliary variable. Pak J Stat 23:1–9

Shabbir J, Gupta S (2010) Estimation of the finite population mean in two phase sampling when auxiliary variables are attributes. Hacet J Math Stat 39:121–129

Singh HP, Kumar S (2010) Improves estimation of population mean under two phase sampling with subsampling the non-respondents. J Stat Plann Inference 140:2536–2550

Singh HP, Solanki RS (2012) Improved estimation of population mean in simple random sampling using information on auxiliary attribute. Appl Math Comput 218:7798–7812

Singh R, Chauhan P, Sawan N, Smarandache F (2008) Ratio estimators in simple random sampling using information on auxiliary attribute. Pak J Stat Oper Res 4:47–53

Sukhatme PV, Sukhatme BV (1970) Sampling theory of surveys with applications. Asia Publishing House, New Dehli