Applied Mathematics and Computation 241 (2014) 14–24

Contents lists available at ScienceDirect

Applied Mathematics and Computation

journal homepage: www.elsevier.com/locate/amc

An improved estimator of finite population mean when using two auxiliary attributes

Abdul Haq, Javid Shabbir *

Department of Statistics, Quaid-i-Azam University, 45320 Islamabad, Pakistan

ARTICLE INFO

Keywords:
Auxiliary attribute
Bias
Mean squared error
Point bi-serial correlation
Simple random sampling
Two-phase sampling

ABSTRACT

In this paper, we propose an improved estimator of finite population mean using information on two auxiliary attributes under simple random sampling (SRS) and two-phase sampling schemes. The bias and mean squared error (MSE) expressions of the proposed estimators are obtained up to first order of approximation. It is shown theoretically, that the proposed estimator is more efficient than the ratio-product type estimator, regression estimator, difference type estimator and Malik and Singh (2013) [13] estimator in both SRS and two phase sampling. An empirical study using two data sets is also conducted to support the theoretical findings.

1. Introduction

In survey sampling, the auxiliary information that is correlated with the study variable is frequently used to increase the precision of the estimator. This auxiliary information can be quantified in form of the auxiliary variables and attributes. For this reason, several authors have exploited use of the auxiliary variables and attributes at the estimation stage to increase efficiency of the estimator. For example, the diameter of a tree can be used as a key auxiliary variable when estimating the average height of trees in a forest. Similarly, the breed of a cow is an important auxiliary attribute when estimating average milk yield. Moreover, to estimate the mean hourly wages earned by the people, the auxiliary information can be used in form of the education, marital status, the region of residence, etc. In these examples, the point bi-serial correlation between the study variable and the auxiliary attribute exists and can lead to more precise estimates. For more details, see Naik and Gupta [16], Jhajj et al. [8], Shabbir and Gupta [18,19] and references cited therein.

In literature several authors have suggested efficient estimators of finite population mean using information on the auxiliary variables and the auxiliary attributes. Some relevant references include Dalabehara and Sahoo [2,3], Tracy et al. [25], Kadilar and Cingi [9], Gupta and Shabbir [6], Singh and Vishwakarma [23], Singh et al. [24], Koyuncu and Kadilar [12], Shabbir and Gupta [19], Abd-Elfattah et al. [1], Grover and Kaur [5], Koyuncu [10,11], Malik and Singh [13,14], Singh and Malik [15], Singh and Solanki [22], and Haq and Shabbir [7]. In this paper, we use the auxiliary information in form of the auxiliary attribute that is correlated with the study variable.

Consider a finite population $U = \{U_1, U_2, \ldots, U_N\}$ of size N. Assume that there is a complete dichotomy in the population depending on the presence and absence of the auxiliary attribute ϕ_j, for $j = 1, 2$. Let y_i, ϕ_{i1} and ϕ_{i2} be the observations of the study variable and the two auxiliary attributes associated with the ith unit ($i = 1, 2, \ldots, N$). Let $\phi_{ij} = 1$, if the ith unit in the

* Corresponding author.
E-mail addresses: aaabdulhaq@gmail.com (A. Haq), javidshabbir@gmail.com (J. Shabbir).

http://dx.doi.org/10.1016/j.amc.2014.04.069

population possesses the auxiliary attribute ϕ_j, and $\phi_{ij} = 0$ otherwise. A random sample of size n is drawn from U by using simple random sampling (SRS) without replacement (SRSWOR). Let $A_j = \sum_{i=1}^{N} \phi_{ij}$ and $a_j = \sum_{i=1}^{n} \phi_{ij}$ be the total number of units in the population and sample respectively that possess an auxiliary attribute ϕ_j $(j = 1, 2)$. Similarly, let the corresponding population and sample proportions respectively are $P_j = \frac{1}{N}\sum_{i=1}^{N} \phi_{ij} = \frac{A_j}{N}$ and $p_j = \frac{1}{n}\sum_{i=1}^{n} \phi_{ij} = \frac{a_j}{n}$, for $j = 1, 2$. Let $s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$, $s_{\phi_j}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\phi_{ij} - p_j)^2$ and $S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{Y})^2$, $S_{\phi_j}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(\phi_{ij} - P_j)^2$ respectively be the sample and population variances of the study variable (y) and the auxiliary attribute (ϕ_j), where $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ and $\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i$. Let $s_{y\phi_j} = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})(\phi_{ij} - p_j)$ and $S_{y\phi_j} = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{Y})(\phi_{ij} - P_j)$ be the sample and population point bi-serial covariance between y and ϕ_j, respectively. Similarly, let $\hat{\rho}_{y\phi_j} = \frac{s_{y\phi_j}}{s_y s_{\phi_j}}$ and $\rho_{y\phi_j} = \frac{S_{y\phi_j}}{S_y S_{\phi_j}}$ be the sample and population bi-serial correlation coefficient between y and ϕ_j, respectively. Let $s_{\phi_1\phi_2} = \frac{1}{n-1}\sum_{i=1}^{n}(\phi_{i1} - p_1)(\phi_{i2} - p_2)$ and $\hat{\rho}_{\phi_1\phi_2} = \frac{s_{\phi_1\phi_2}}{s_{\phi_1} s_{\phi_2}}$ be the sample phi-covariance and phi-correlation between the two auxiliary attributes ϕ_1 and ϕ_2, respectively. Similarly, the corresponding population phi-covariance and phi-correlation between ϕ_1 and ϕ_2 are $S_{\phi_1\phi_2} = \frac{1}{N-1}\sum_{i=1}^{N}(\phi_{i1} - P_1)(\phi_{i2} - P_2)$ and $\rho_{\phi_1\phi_2} = \frac{S_{\phi_1\phi_2}}{S_{\phi_1} S_{\phi_2}}$, respectively. Let $C_y = \frac{S_y}{\bar{Y}}$ and $C_{\phi_j} = \frac{S_{\phi_j}}{P_j}$ be the coefficients of variation of y and ϕ_j, respectively.

In order to obtain the bias and mean squared error (MSE) of the estimators, we define the following relative error terms. Let $\xi_0 = \frac{\bar{y} - \bar{Y}}{\bar{Y}}$ and $\xi_j = \frac{p_j - P_j}{P_j}$, such that $E(\xi_0) = E(\xi_j) = 0$ for $j = 1, 2$; $E(\xi_0^2) = \lambda_1 C_y^2$, $E(\xi_1^2) = \lambda_1 C_{y\phi_1}^2$, $E(\xi_2^2) = \lambda_1 C_{y\phi_2}^2$, $E(\xi_0\xi_1) = \lambda_1 \rho_{y\phi_1} C_y C_{\phi_1}$, $E(\xi_0\xi_2) = \lambda_1 \rho_{y\phi_2} C_y C_{\phi_2}$, $E(\xi_1\xi_2) = \lambda_1 \rho_{\phi_1\phi_2} C_{\phi_1} C_{\phi_2}$, where $\lambda_1 = (1/n) - (1/N)$.

The outline of the paper is as follows: In Section 2, we overview several existing estimators of $\bar{Y}$ in SRS. In Section 3, an improved estimator of $\bar{Y}$ using information on two auxiliary attributes is proposed. The expressions for bias and MSE are obtained under first order of approximation. In Section 4, the existing and proposed estimators are considered under two-phase sampling. In Section 5, we provide theoretical comparison to evaluate the performances of the estimators under both sampling schemes considered here. An empirical study is conducted in Section 6, and concluding remarks are given in Section 7.

2. Existing estimators in SRS

In this section, we consider several estimators of finite population mean.

2.1. Regression estimator

The regression estimator of $\bar{Y}$ based on two auxiliary attributes, is given by

$$\hat{\bar{Y}}_{REG} = \bar{y} + b_1(P_1 - p_1) + b_2(P_2 - p_2), \tag{2.1}$$

where $b_j = \frac{s_{y\phi_j}}{s_{\phi_j}^2}$ for $j = 1, 2$.

The MSE of $\hat{\bar{Y}}_{\mathrm{REG}}$, up to first order of approximation, is given by

$$\mathrm{MSE}\left(\hat{\bar{Y}}_{\mathrm{REG}}\right) \cong \lambda_1 \bar{Y}^2 C_y^2 \left(1 - \rho_{y\phi_1}^2 - \rho_{y\phi_2}^2 + 2\rho_{y\phi_1}\rho_{y\phi_2}\rho_{\phi_1\phi_2}\right). \tag{2.2}$$

After some simplifications, (2.2) can be written as

$$\mathrm{MSE}\left(\hat{\bar{Y}}_{\mathrm{REG}}\right) \cong \lambda_1 \bar{Y}^2 C_y^2 \left\{\rho_{\phi_1\phi_2}^2 + \left(1 - \rho_{\phi_1\phi_2}^2\right)\left(1 - R_{y\cdot\phi_1\phi_2}^2\right)\right\}, \tag{2.3}$$

where $R_{y\cdot\phi_1\phi_2}^2 = \frac{\rho_{y\phi_1}^2 + \rho_{y\phi_2}^2 - 2\rho_{y\phi_1}\rho_{y\phi_2}\rho_{\phi_1\phi_2}}{1 - \rho_{\phi_1\phi_2}^2}$ is the multiple correlation coefficient of y on ϕ_1 and ϕ_2.

2.2. Malik and Singh [13] estimator

Recently Malik and Singh [13] has suggested an improved estimator of $\bar{Y}$ using two auxiliary attributes, given by

$$\hat{\bar{Y}}_{\mathrm{MS}} = \bar{y}\exp\left(\frac{P_1 - p_1}{P_1 + p_1}\right)^{\gamma_1}\exp\left(\frac{P_2 - p_2}{P_2 + p_2}\right)^{\gamma_2} + b_1(P_1 - p_1) + b_2(P_2 - p_2), \tag{2.4}$$

where γ_1 and γ_2 are two unknown constants.
The MSE of $\hat{\bar{Y}}_{\mathrm{MS}}$, up to first order of approximation, is given by

$$\mathrm{MSE}\left(\hat{\bar{Y}}_{\mathrm{MS}}\right) \cong \frac{1}{4}\lambda_1\Big[4P_1^2\beta_1^2 C_{\phi_1}^2 + 4Y^2 C_y^2 + \bar{Y}^2 C_{\phi_1}^2\gamma_1^2 + C_{\phi_2}(2P_2\beta_2 + \bar{Y}\gamma_2)\left\{C_{\phi_2}(2P_2\beta_2 + \bar{Y}\gamma_2) + 2\bar{Y}C_{\phi_1}\gamma_1\rho_{\phi_1\phi_2}\right\}$$
$$+ 4P_1\beta_1 C_{\phi_1}\left\{\bar{Y}C_{\phi_1}\gamma_1 + C_{\phi_2}(2P_2\beta_2 + \bar{Y}\gamma_2)\rho_{\phi_1\phi_2} - 2\bar{Y}C_y\rho_{y\phi_1}\right\} - 4\bar{Y}C_y\left\{\bar{Y}C_{\phi_1}\gamma_1\rho_{y\phi_1} + C_{\phi_2}(2P_2\beta_2 + \bar{Y}\gamma_2)\rho_{y\phi_2}\right\}\Big],$$

where $\beta_j = \frac{S_{y\phi_j}}{S_{\phi_j}^2}$ for $j = 1, 2$.