World Applied Sciences Journal 7 (11): 1419-1422, 2009
ISSN 1818-4952

On a New Family of Estimators using Multiple Auxiliary Attributes

[1]*Muhammad Hanif,* [2]*Inam ul Haq and* [3]*Muhammad Qaiser Shahbaz*

[1]Lahore University of Management Sciences, Lahore, Pakistan
[2]National College of Business Administration and Economics, Lahore, Pakistan
[3]Department of Mathematics, COMSATS Institute of Information Technology, Lahore, Pakistan

Abstract: A general family of estimators has been proposed and general expression of mean square error of these estimators has been derived by [1]. In this paper we have proposed a generalized family of estimators based on the information of "*k*" auxiliary attributes. Three different cases have been discussed that include the full, partial and no information cases. The family has been proposed for single-phase sampling in case of full information and for two-phase sampling in case of partial and no information cases. The expression for mean square error has been derived in all three cases. It is found that the proposed family has smaller mean square error than given in [1].

Key words: Multi-auxiliary attributes . generalization . dichotomy . efficiency . bi-serial correlation coefficient . full . partial and no information . multiple bi-serial correlation coefficient

INTRODUCTION

A family of estimators using single auxiliary attribute has been introduced by [1]. In this paper, we have proposed a new class of estimator by using information on "k" auxiliary attributes. The new class of estimators is a general extension of the class of estimators proposed by [1]. For this let $(y_i, \tau_{i1}, \tau_{i2},\ldots,\tau_{ik})$ be the ith sample point from a population of size N, where t_j (j=1,2,…,k) is the value of jth auxiliary attribute. We suppose that the complete dichotomy is recorded for each attribute so that $\tau_{ij} = 1$ if ith unit of population possesses jth attribute and t_j and 0 otherwise. Let

$$A_j = \sum_{i=1}^{N} \tau_{ij}$$

and

$$a_j = \sum_{i=1}^{n} \tau_{ij}$$

be the total number of units in the population and sample respectively, possessing attribute τ_j. Let $P_j = N^{-1}A_j$ and $p_j = n^{-1}a_j$ be the corresponding proportion of units possessing attributes τ_j. Let us define $\bar{e}_y = \bar{y} - \bar{Y}$ and $\bar{e}_{\tau_j} = p_j - P_j$ with following properties:

$$E(\bar{e}_y^2) = \theta S_y^2$$

$$E(\bar{e}_{\tau_j}^2) = \theta S_{\tau_j}^2$$

$$E(\bar{e}_y \bar{e}_{\tau_j}) = \theta S_y S_{\tau_j} \rho_{pb_j}$$

$$E\left(\bar{e}_{\tau_j} \bar{e}_{\tau_\psi}\right) = \theta S_{\tau_j} S_{\tau_\psi} Q_{j\psi} \;\&\; j \neq \psi$$

$$E(\bar{e}_y) = 0 = E(\bar{e}_{\tau_j})$$

where $\theta = n^{-1}\text{-}N^{-1}$ and

$$S_{y\tau_j} = \frac{1}{N-1}\sum_{j=1}^{N}\left(y_i - \bar{Y}\right)\left(\tau_{ij} - P_j\right)$$

Suppose further that $\rho_{pbj} = S_{y\tau j}/(S_y S_{\tau j})$ be the point bi-serial correlation coefficient and $Q_{j\psi}$ $(-1 \le Q_{j\psi} \le +1)$ is coefficient of association. Let n_1 and n_2 be the size of first-phase and second-phase sample respectively, so that $n_2 < n_1$ and $p_{j(1)}$, $p_{j(2)}$ are proportion of units possessing attribute t_j in first-phase and second-phase sample respectively. The mean of main variable of interest at second phase is denoted by $\bar{y}_2$.

NOTATIONS

We define following notations for use in deriving the mean square error:

Corresponding Author: Dr. Muhammad Hanif, Lahore University of Management Sciences, Lahore, Pakistan

$$\bar{e}_{y_2} = \bar{y}_2 - \bar{Y}$$

$$\bar{e}_{\tau_j(1)} = p_{j(1)} - P_j$$

$$\bar{e}_{\tau_j(2)} = p_{j(2)} - P_j$$

$$(j=1,2\ldots k), \theta_3 = \theta_2 - \theta_1.$$

We also define following expectations:

$$E\left(\bar{e}_{y_2}\right) = \theta_2 S_y^2$$

$$E\left(\bar{e}_{\tau_{j(1)}} - \bar{e}_{\tau_{j(2)}}\right)^2 = \theta_3 S_{\tau_j}^2$$

$$E\left\{\bar{e}_{y_2}\left(\bar{e}_{\tau_{j(2)}} - \bar{e}_{\tau_{j(1)}}\right)\right\} = \theta_3 S_y S_{\tau_j} \rho_{pbj},$$

$$E\left[\left(\bar{e}_{\tau_{j(2)}} - \bar{e}_{\tau_{j(1)}}\right)\left(\bar{e}_{\tau_{\psi(2)}} - \bar{e}_{\tau_{\psi(1)}}\right)\right]$$
$$= \theta_3 S_{\tau_j} S_{\tau_\psi} Q_{j\psi}, \quad s'_{y\tau} S_\tau^{-1} s_{y\tau} = \left(\rho^2_{y.\tau_1\tau_2\ldots\tau_k}\right) S_y^2$$

$$s'_{y\tau_1} S_{\tau_1}^{-1} s_{y\tau_1} = \left(\rho^2_{y.\tau_1\tau_2\ldots\tau_m}\right) S_y^2$$

$$s'_{y\tau_2} S_{\tau_2}^{-1} s_{y\tau_2} = \left(\rho^2_{y.\tau_{m+1}\tau_{m+2}\ldots\tau_k}\right) S_y^2$$

where $\rho^2_{y.\tau_1\tau_2\ldots\tau_k}$ is squared multiple bi-serial correlation coefficient. These notation will be used in developing the MSE of the new family of estimators.

SOME PREVIOUS ESTIMATORS BASED ON AUXILIARY ATTRIBUTES

Single-phase sampling (Full information case): (i) If information on a single auxiliary attribute t_1 is known then a family of estimator suggested by [1] is given as

$$T_{1(1)} = g_\omega\left(\bar{y}, v_1\right) \tag{3.1}$$

where $v_1 = p_1/P_1$ and $g_\omega\left(\bar{y}, v_1\right)$ is a parametric function of $\bar{y}$ and v_1 such that $g_\omega\left(\bar{Y}, 1\right) = \bar{Y}$ and satisfy certain regularity conditions. The mean square error of (2.1) is:

$$MSE\left(T_{1(1)}\right) \approx \theta S_y^2\left(1 - \rho^2_{pb1}\right) \tag{3.2}$$

where ρ^2_{pb1} is squared point bi-serial correlation coefficient.

Two-phase sampling (No information case): A family of estimator for two phase sampling by [1] is given as

$$T_{3(2)} = g_\omega\left(\bar{y}_2, v_{1d}\right) \tag{3.3}$$

where $v_{1d} = p_{1(2)}/P_{1(1)}$, such that $g_\omega\left(\bar{Y}, 1\right) = \bar{Y}$.
The mean square error of (3.3) is:

$$MSE\left(T_{3(2)}\right) \approx S_y^2\left(\theta_2 - \theta_3 \rho^2_{pb1}\right) \tag{3.4}$$

In the following section we develop the general family of estimators by using information on "k" auxiliary attributes.

NEW FAMILY OF ESTIMATORS.

Generalized estimator using "k" auxiliary attributes for full information case: Suppose that population proportion P_j is known for all the auxiliary attributes. Using this full information we propose a general family of estimators as:

$$T_{3(1)} = g_\omega\left(\bar{y}, v_1, v_2, \ldots\ldots\ldots, v_k\right) \tag{4.1}$$

where $v_j = p_j/P_j$, $v_j>0$ and p_j is the sample proportions of jth attributes. Also $g_\omega\left(\bar{y}, v_1, v_2, \ldots\ldots\ldots, v_k\right)$ is the parametric function such that $g_\omega\left(\bar{Y}, 1, 1, \ldots.., 1\right) = \bar{Y}$ and the point $\left(\bar{y}, v_1, v_2, \ldots.. v_k\right)$ are to be in a bounded set in R_k containing a point $\left(\bar{Y}, 1, 1, \ldots.., 1\right)$. The attributes τ_j are significantly correlated with main variable. We consider the following estimator of the family defined in equation (4.1):

$$t_{3(1)} = \bar{y} + \sum_{j=1}^{k} \alpha_j\left(v_j - 1\right) = \bar{y} + a'\left(v - 1\right) = \bar{y} + \alpha' f \tag{4.2}$$

where

$$a_{(k \times 1)} = \left[\alpha_j\right], v_{(k \times 1)} = \left[v_j\right], f = v - 1$$

and 1 is a vector of one's. Using $\bar{y} = \bar{Y} + \bar{e}_y$ in (4.2), squaring and applying expectation we have:

$$MSE\left(t_{3(1)}\right) = \theta S_y^2 + \theta a' S_t a + 2\theta a' s_{y\tau} \tag{4.3}$$

where $\theta = n^{-1} - N^{-1}$

$$E\left(f f'\right) = \theta S_\tau$$

is the covariance matrix of φ and $E\left(f\,\bar{e}_y\right) = \theta s_{y\tau}$ is the vector of covariance between Y and φ. Partially differentiating (4.3) with respect to α and equating the derivative to zero we have $a = -S_\tau^{-1}s_{y\tau}$. Further, by using the value of α in (4.3), the mean square error of (4.2) is given as:

$$MSE\left(t_{3(1)}\right) = \theta S_y^2\left(1 - \rho^2_{y.\tau_1\tau_2\ldots\tau_k}\right) \qquad (4.4)$$

where $\rho^2_{y.\tau_1\tau_2\ldots\tau_k}$ is defined earlier. Comparing (4.4) with (3.2) we can readily see that MSE $(t_{3(1)})$<MSE$(t_{1(1)})$
.

Generalized estimator using "k" auxiliary attributes for no information case: We propose a family of estimators for two-phase sampling when information of auxiliary attributes is not known for the population. We propose the following family of estimators:

$$T_{4(2)} = g_\omega\left(\bar{y}_2, v_{1d}, v_{2d}, \ldots\ldots\ldots\ldots, v_{kd}\right) \qquad (4.5)$$

where

$$v_{jd} = p_{j(2)}/p_{j(1)};\; v_{jd} > 0$$

Under the conditions; stated for (3.1); we consider the following estimator of the family defined in (4.5):

$$t_{4(2)} = \bar{y}_2 + \sum_{j=1}^{k}\alpha_j\left(v_{jd} - 1\right) = \bar{y}_2 + a^{/}\left(v_d - 1\right) = \bar{y}_2 + a^{/}f_d \qquad (4.6)$$

where φ_d = v_d-1. Now, using $\bar{y}_2 = \bar{Y} + \bar{e}_{y_2}$ in (4.6) we have

$$t_{4(2)} - \bar{Y} = \bar{e}_{y_2} + a^{/}f_d$$

Squaring and applying expectation we have:

$$MSE\left(t_{4(2)}\right) = \theta_2 S_y^2 + \left(\theta_2 - \theta_1\right)a^{/}S_\tau a - 2\left(\theta_2 - \theta_1\right)a^{/}s_{y\tau} \qquad (4.7)$$

Partially differentiating (4.7) w.r.t α and equating to zero we have $a = -S_\tau^{-1}s_{y\tau}$. Further, by using this value of α is (4.7) and simplifying, the mean square error of $t_{4(2)}$ is:

$$MSE\left(t_{4(2)}\right) = S_y^2\left\{\theta_2\left(1 - \rho^2_{y.\tau_1\tau_2\ldots\tau_k}\right) + \theta_1\rho^2_{y.\tau_1\tau_2\ldots\tau_k}\right\} \qquad (4.8)$$

where $\rho^2_{y.\tau_1\tau_2\ldots\tau_k}$ is defined earlier.

Generalized estimator for "k" auxiliary attributes (With "m" known and "m<k") for partial information case: Suppose that population proportion P_j are known for j= (1,2…m) auxiliary attributes and the population proportion P_j is unknown for j= (m+1 m+2…k) attributes. Using such partial information we propose following general family of estimators:

$$T_{5(2)} = g_\omega\left(\begin{matrix}\bar{y}_2, v_1, v_2, \ldots v_m, v_{(m+1)d}, \\ v_{(m+2)d}, \ldots, v_{kd}\end{matrix}\right) \qquad (4.9)$$

where

$$v_j = p_{j(1)}/P_j,\ (j = 1,2\ldots m); v_{jd} = p_{j(2)}/p_{j(1)},\ (j = m+1, m+2\ldots k), v_j \,\&\, v_{jd} > 0.$$

Under the conditions as of (4.1) we consider the following estimator of the family defined in equation (4.6),

$$t_{5(2)} = \bar{y}_2 + \sum_{j=1}^{m}\alpha_j\left(v_j - 1\right) + \sum_{j=m+1}^{k}\alpha_j\left(v_{jd} - 1\right) = \bar{y}_2 + a_1^{/}\left(v_1 - 1\right) + a_2^{/}\left(v_2 - 1\right)$$

$$= \bar{y}_2 + a_1^{/}f_1 + a_2^{/}f_2 \qquad (4.10)$$

Squaring and applying expectation we have:

$$MSE\left(t_{5(2)}\right) = \theta_2 S_y^2 + \theta_1 a_1^{/}S_{\tau_1}a_1 + \left(\theta_2 - \theta_1\right)a_2^{/}S_{\tau_2}a_2 + 2\theta_1 a_1^{/}s_{y\tau_1} + 2\left(\theta_2 - \theta_1\right)a_2^{/}s_{y\tau_2} \qquad (4.11)$$

Partially differentiating (3.11) w.r.t. α_1 and α_2; equating the derivatives to zero and solving we have the following optimum values of α_1 and α_2:

$$a_1 = -S_{\tau_1}^{-1}s_{y\tau_1} \qquad (4.12)$$

$$a_2 = -S_{\tau_2}^{-1}s_{y\tau_2} \qquad (4.13)$$

Using (4.12) and (4.13) in (4.11), the minimum mean square error of $t_{5(2)}$ is given as:

$$MSE\left(t_{5(2)}\right)=S_y^2\left\{\begin{array}{l}\theta_2\left(1-\rho^2_{y.\tau_{m+1}\tau_{m+2}\cdots\tau_k}\right)\\+\theta_1\left(\rho^2_{y.\tau_{m+1}\tau_{m+2}\cdots\tau_k}-\rho^2_{y.\tau_1\tau_2\cdots\tau_m}\right)\end{array}\right\} \quad (4.14)$$

where $\rho^2_{y.\tau_{m+1}\tau_{m+2}\cdots\tau_k}$ and $\rho^2_{y.\tau_1\tau_2\cdots\tau_m}$ is the squared multiple bi-serial correlation coefficient.

RESULTS AND DISCUSSION

The information on *k* auxiliary attributes has been utilized to develop the generalized family of estimators for single and two phase sampling. There could be number of estimators of families proposed in (4.1), (4.5) and (4.9), the special members have been given in (4.2), (4.7) and (4.13). The expression for mean square error of the resulting estimators has been given in (4.5), (4.11) and (4.15). It can be easily seen that the mean square errors given in (4.5), (4.11) and (4.15) are smaller as compared with the expression given by [1].

The optimum value of α_j involve some population parameters, which are assumed to be known for the efficient use of proposed families $T_{3(1)}$, $T_{4(2)}$ and $T_{5(2)}$. In case these parameters are unknown, these can be estimated from the sample. If we follow approach of [2], the estimator of proposed families $T_{3(1)}$, $T_{4(2)}$ and $T_{5(2)}$ will have the same minimum mean square, if we replace the unknown value of parameters involved in optimum value of α_j with their consistent estimators.

REFERENCES

1. Jhajj, H.S., M.K. Sharma and L.K. Grover, 2006. A family of estimators of population mean using information on auxiliary attribute. Pakistan Journal of Statistics, 22 (1): 43-50.
2. Srivastava, S.K. and H.S. Jhajj, 1983. A class of estimators of the population means using multi-auxiliary information. Calcutta Statistics Association Bulletin, 32: 47-56.