# A Generalized Family Of Estimators For Estimating Population Mean Using Two Auxiliary Attributes

**[1]Sachin Malik, [†1]Rajesh Singh and [2]Florentin Smarandache**

[1]Department of Statistics, Banaras Hindu University
Varanasi-221005, India

[2]Chair of Department of Mathematics, University of New Mexico, Gallup, USA

† Corresponding author, rsinghstat@gmail.com

## *Abstract*

This paper deals with the problem of estimating the finite population mean when some information on two auxiliary attributes are available. A class of estimators is defined which includes the estimators recently proposed by Malik and Singh (2012), Naik and Gupta (1996) and Singh et al. (2007) as particular cases. It is shown that the proposed estimator is more efficient than the usual mean estimator and other existing estimators. The study is also extended to two-phase sampling. The results have been illustrated numerically by taking empirical population considered in the literature.

***Keywords*** Simple random sampling, two-phase sampling, auxiliary attribute, point bi-serial correlation, phi correlation, efficiency.

## *1. Introduction*

There are some situations when in place of one auxiliary attribute, we have information on two qualitative variables. For illustration, to estimate the hourly wages we can use the information on marital status and region of residence (see Gujrati and Sangeetha (2007), page-311). Here we assume that both auxiliary attributes have significant point bi-serial correlation with the study variable and there is significant phi-correlation (see Yule (1912)) between the auxiliary attributes. The use of auxiliary information can increase the precision of an estimator when study variable Y is highly correlated with auxiliary variables X. In survey sampling, auxiliary variables are present in form of ratio scale variables (e.g. income, output, prices, costs, height and temperature) but sometimes may present in the form of qualitative or nominal scale such as sex, race, color, religion, nationality and geographical region. For example, female workers are found to earn less than their male counterparts do or non-white workers are found to earn less than whites (see Gujrati and Sangeetha (2007), page 304). Naik and Gupta (1996) introduced a ratio estimator when the study variable and the auxiliary attribute are positively correlated. Jhajj et al. (2006) suggested a family of estimators for the population mean in single and two-phase sampling when the study variable and auxiliary attribute are positively correlated. Shabbir and Gupta (2007), Singh et al. (2008), Singh et al. (2010) and Abd-Elfattah et al. (2010) have considered the problem of estimating population mean $\overline{Y}$ taking into consideration the point biserial correlation between auxiliary attribute and study variable.

## *2. Some Estimators in Literature*

In order to have an estimate of the study variable y, assuming the knowledge of the population proportion P, Naik and Gupta (1996) and Singh et al. (2007) respectively, proposed following estimators:

$$t_1 = \bar{y}\left(\frac{P_1}{p_1}\right) \qquad (2.1)$$

$$t_2 = \bar{y}\left(\frac{p_2}{P_2}\right) \qquad (2.2)$$

$$t_3 = \bar{y}\exp\left(\frac{P_1 - p_1}{P_1 + p_1}\right) \qquad (2.3)$$

$$t_4 = \bar{y}\exp\left(\frac{p_2 - P_2}{p_2 + P_2}\right) \qquad (2.4)$$

The Bias and MSE expression's of the estimator's $t_i$ (i=1, 2, 3, 4) up to the first order of approximation are, respectively, given by

$$B(t_1) = \overline{Y} f_1 C_{p_1}^2 \left[1 - K_{pb_1}\right] \qquad (2.5)$$

$$B(t_2) = \overline{Y} f_1 K_{pb_2} C_{p_2}^2 \qquad (2.6)$$

$$B(t_3) = \overline{Y} f_1 \frac{C_{p_2}^2}{2}\left[\frac{1}{4} - K_{pb_2}\right] \qquad (2.7)$$

$$B(t_4) = \overline{Y} f_1 \frac{C_{p_2}^2}{2}\left[\frac{1}{4} + K_{pb_2}\right] \qquad (2.8)$$

$$MSE(t_1) = \overline{Y}^2 f_1\left[C_y^2 + C_{p_1}^2\left(1 - 2K_{pb_1}\right)\right] \qquad (2.9)$$

$$MSE(t_2) = \overline{Y}^2 f_1\left[C_y^2 + C_{p_1}^2\left(1 + 2K_{pb_2}\right)\right] \qquad (2.10)$$

$$MSE(t_3) = \bar{Y}^2 f_1\left[C_y^2 + C_{p_1}^2\left(\frac{1}{4} - K_{pb_2}\right)\right] \tag{2.11}$$

$$MSE(t_4) = \bar{Y}^2 f_1\left[C_y^2 + C_{p_2}^2\left(\frac{1}{4} + K_{pb_2}\right)\right] \tag{2.12}$$

where, $f_1 = \frac{1}{n} - \frac{1}{N}$ , $S_{\phi_j}^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(\phi_{ji} - P_j\right)^2$, $S_{y\phi_j} = \frac{1}{N-1}\sum_{i=1}^{N}\left(y_i - \bar{Y}\right)\left(\phi_{ji} - P_j\right)$,

$\rho_{pb_j} = \frac{S_{y\phi_j}}{S_y S_{\phi_j}}$, $C_y = \frac{S_y}{\bar{Y}}$, $C_{pj} = \frac{S_{\phi_j}}{P_j}$; $(j = 1,2)$,

$K_{pb_1} = \rho_{pb_1}\frac{C_y}{C_{p_1}}$, $K_{pb_2} = \rho_{pb_2}\frac{C_y}{C_{p_2}}$.

$s_{\phi_1\phi_2} = \frac{1}{n-1}\sum_{i=1}^{n}(\phi_{1i} - p_1)(\phi_{2i} - p_2)$ and $\rho_\phi = \frac{s_{\phi_1\phi_2}}{s_{\phi_1}s_{\phi_2}}$ be the sample phi-covariance and phi-correlation between $\phi_1$ and $\phi_2$ respectively, corresponding to the population phi-covariance and phi-correlation $S_{\phi_1\phi_2} = \frac{1}{N-1}\sum_{i=1}^{N}(\phi_{1i} - P_1)(\phi_{2i} - P_2)$

and $\rho_\phi = \frac{S_{\phi_1\phi_2}}{S_{\phi_1}S_{\phi_2}}$.

Malik and Singh (2012) proposed estimators $t_5$ and $t_6$ as

$$t_5 = \bar{y}\left(\frac{P_1}{p_1}\right)^{\alpha_1}\left(\frac{P_2}{p_2}\right)^{\alpha_2} \tag{2.13}$$

$$t_6 = \bar{y}\exp\left(\frac{P_1 - p_1}{P_1 + p_1}\right)^{\beta_1}\exp\left(\frac{p_2 - P_2}{p_2 + P_2}\right)^{\beta_2} \tag{2.14}$$

where $\alpha_1, \alpha_2, \beta_1$ and $\beta_2$ are real constants.

The Bias and MSE expression's of the estimator's $t_5$ and $t_6$ up to the first order of approximation are, respectively, given by

$$B(t_5) = \bar{Y}f_1\left(C_{p_1}^2\left[\frac{\alpha_1^2}{2} + \frac{\alpha_1}{2} - \alpha_1 k_{pb_1}\right] + C_{p_2}^2\left[\frac{\alpha_2^2}{2} + \frac{\alpha_2}{2} - \alpha_2 k_{pb_2} + \alpha_1\alpha_2 k_\phi\right]\right) \tag{2.15}$$

$$B(t_6) = \bar{Y} f_1 \left[ C_{p_1}^2 \left( \frac{\beta_1^2}{4} - \frac{\beta_1}{2} K_{pb_1} \right) + C_{p_2}^2 \left( \frac{\beta_2^2}{4} + \frac{\beta_2}{2} K_{pb_2} - \frac{\beta_1 \beta_2}{4} K_{\phi} \right) \right] \tag{2.16}$$

$$MSE(t_5) = \bar{Y}^2 f_1 \left[ C_y^2 + C_{p_1}^2 \left( \alpha_1^2 - 2\alpha_1 K_{pb_1} \right) + C_{p_2}^2 \left( \alpha_2^2 - 2\alpha_2 K_{pb_2} + 2\alpha_1 \alpha_2 K_{\phi} \right) \right] \tag{2.17}$$

$$MSE(t_6) = \bar{Y}^2 f_1 \left[ C_y^2 + C_{p_1}^2 \left( \frac{\beta_1^2}{4} - \beta_1 K_{pb_1} \right) + C_{p_2}^2 \left( \frac{\beta_2^2}{4} - \frac{\beta_1 \beta_2}{2} K_{\varphi} + \beta_2 K_{pb_1} \right) \right] \tag{2.18}$$

## *3. The Suggested Class of Estimators*

Using linear combination of $t_i \left( i = 0,1,2 \right)$, we define an estimator of the form

$$t_p = \sum_{i=0}^{3} w_i t_i \in H \tag{3.1}$$

Such that, $\sum_{i=0}^{3} w_i = 1$ and $w_i \in R$ (3.2)

Where,

$$t_0 = \bar{y}, \; t_1 = \bar{y} \left[ \frac{L_1 P_1 + L_2}{L_1 p_1 + L_2} \right]^{\alpha_1} \left[ \frac{L_3 P_2 + L_4}{L_3 p_2 + L_4} \right]^{\alpha_2}$$

$$\text{and } t_2 = \exp \left[ \frac{(L_5 P_1 + L_6) - (Lp_1 + L_6)}{(L_1 P_1 + L_2) + (L_5 p_1 + L_6)} \right]^{\beta_1} \exp \left[ \frac{(L_7 p_2 + L_6) - (L_7 P_2 + L_8)}{(L_7 p_2 + L_2) + (L_7 P_2 + L_8)} \right]^{\beta_2}$$

where $w_i \left( i = 0,1,2 \right)$ denotes the constants used for reducing the bias in the class of estimators, H denotes the set of those estimators that can be constructed from $t_i \left( i = 0,1,2 \right)$ and R denotes the set of real numbers (for detail see Singh et. al (2008)). Also, $L_i \left( i = 1,2,...,8 \right)$ are either real numbers or the functions of the known parameters of the auxiliary attributes.

Expressing $t_p$ in terms of e's, we have

$$t_p = \bar{Y} \left( 1 + e_0 \right) \left[ \begin{array}{l} w_0 + w_1 \left( 1 + \varphi_1 e_1 \right)^{-\alpha_1} \left( 1 + \varphi_2 e_2 \right)^{-\alpha_2} \\ + w_2 \exp \left( -\theta_1 e_1 \left[ 1 + \theta_1 e_1 \right]^{-1} \right)^{\beta_1} \\ \exp \left( -\theta_2 e_2 \left[ 1 + \theta_2 e_2 \right]^{-1} \right)^{\beta_2} \end{array} \right] \tag{3.3}$$

where,

$$\left.\begin{aligned} \varphi_1 &= \frac{L_1P_1}{L_1P_1 + L_2} \\ \varphi_2 &= \frac{L_3P_2}{L_3P_1 + L_4} \\ \theta_1 &= \frac{L_5P_1}{2[L_5P_2 + L_6]} \\ \theta_2 &= \frac{L_7P_2}{2[L_7P_2 + L_8]} \end{aligned}\right\}$$

After expanding, Subtracting $\overline{Y}$ from both sides of the equation (3.3) and neglecting the term having power greater than two, we have

$$\left(t_p - \overline{Y}\right) = \overline{Y}\left[e_0 - w_1\left(\alpha_1\varphi_1 e_1 + \alpha_2\varphi_2 e_2\right) - w_2\left(\beta_1\theta_1 e_1 - \beta_2\theta_2 e_2\right)\right] \tag{3.4}$$

Squaring both sides of (3.4) and then taking expectations, we get MSE of the estimator $t_p$ up to the first order of approximation, as

$$MSE\left(t_p\right) = \overline{Y}^2 f\left[w_1^2T_1 + w_2^2T_2 + 2w_1w_2T_3 - 2w_1T_4 - 2w_2T_5\right] \tag{3.5}$$

where,

$$\left.\begin{aligned} w_1 &= \frac{L_2L_4 - L_3L_5}{L_1L_2 - L_3^2} \\ w_2 &= \frac{L_1L_5 - L_3L_4}{L_1L_2 - L_3^2} \end{aligned}\right\} \tag{3.6}$$

and

$$\left.\begin{aligned} L_1 &= \varphi_1^2\alpha_1^2C_{p_1}^2 + \varphi_2^2\alpha_2^2C_{p_2}^2 + 2\alpha_1\alpha_2\varphi_1\varphi_2 k_\varphi C_{p_2}^2 \\ L_2 &= \theta_1^2\beta_1^2c_{p_1}^2 + \theta_2^2\beta_2^2c_{p_1}^2 - 2\beta_1\beta_2\varphi_1\theta_2 k_\varphi C_{p_2}^2 \\ L_3 &= \alpha 1\beta_1\theta_1C_{p_1}^2 - \alpha_2\beta_2\theta_2C_{p_2}^2 + \alpha_2\beta_1\varphi_2\theta_1 k_\varphi C_{p_2}^2 - \alpha_1\varphi_1\theta_2\beta_2 k_\varphi C_{p_2}^2 \\ L_4 &= \alpha_1\varphi_1 k_{pb_1}C_{p_1}^2 + \alpha_2\varphi_2 k_{pb_2}C_{p_2}^2 \\ L_5 &= \beta_1\theta_1 k_{pb_1}C_{p_1}^2 - \beta_2\theta_2 k_{pb_2}C_{p_2}^2 \end{aligned}\right\} \tag{3.7}$$

## *4. Empirical Study*

Data: (Source: Government of Pakistan (2004))

The population consists rice cultivation areas in 73 districts of Pakistan. The variables are defined as:

Y= rice production (in 000' tonnes, with one tonne = 0.984 ton) during 2003,

$P_1$ = production of farms where rice production is more than 20 tonnes during the year 2002, and

$P_2$ = proportion of farms with rice cultivation area more than 20 ha during the year 2003.

For this data, we have

$N=73,\ \bar{Y}=61.3,\ P_1=0.4247,\ P_2=0.3425,\ S_y^2=12371.4,\ S_{\phi_1}^2=0.225490,\ S_{\phi_2}^2=0.228311,$

$\rho_{pb_1}=0.621,\ \rho_{pb_2}=0.673,\ \rho_{\phi}=0.889.$

**Table 4.1: PRE of different estimators of $\bar{Y}$ with respect to $\bar{y}$.**

| **CHOICE OF SCALERS, when** $w_0=0\ w_1=1\ w_2=0$ | | | | | | |
|---|---|---|---|---|---|---|
| $\alpha_1$ | $\alpha_2$ | $L_1$ | $L_2$ | $L_3$ | $L_4$ | PRE'S |
| 0 | 1 | | | 1 | 0 | 179.77 |
| 1 | 0 | 1 | 0 | | | 162.68 |
| 1 | 1 | 1 | 1 | 1 | 1 | 156.28 |
| -1 | 1 | 1 | 0 | 1 | 0 | 112.97 |
| 1 | 1 | $C_{p_1}$ | $\rho_{pb_1}$ | $C_{p_2}$ | $\rho_{pb_2}$ | 178.10 |
| 1 | 1 | $NP_1$ | $K_{pb_1}$ | $NP_2$ | $K_{pb2}$ | 110.95 |
| -1 | 1 | $NP_1$ | f | $NP_2$ | f | 112.78 |
| -1 | 1 | N | $K_{pb_1}$ | N | $K_{pb2}$ | 112.68 |
| -1 | 1 | $NP_1$ | $P_1$ | $NP_2$ | $P_2$ | 112.32 |
| 1 | 1 | n | $P_1$ | n | $P_2$ | 115.32 |
| -1 | 1 | N | $\rho_{pb_1}$ | N | $\rho_{pb_2}$ | 112.38 |
| -1 | 1 | n | $P_1$ | n | $P_2$ | 113.00 |
| -1 | 1 | N | $P_1$ | N | $P_2$ | 112.94 |
| **When,** $w_0=0\ w_1=0\ w_2=1$ | | | | | | |

| $\beta_1$ | $\beta_2$ | $L_5$ | $L_6$ | $L_7$ | $L_8$ | PRE'S |
|---|---|---|---|---|---|---|
| 1 | 0 | 1 | 0 | 1 | 0 | 141.81 |
| 0 | 1 | 1 | 0 | 1 | 0 | 60.05 |
| 1 | -1 | 1 | 0 | 1 | 0 | 180.50 |
| 1 | -1 | 1 | 1 | 1 | 1 | 127.39 |
| 1 | -1 | 1 | 1 | 1 | 0 | 170.59 |
| 1 | -1 | $C_{p_1}$ | $\rho_{pb_1}$ | $C_{p_2}$ | $\rho_{pb_2}$ | 143.83 |
| 1 | -1 | $NP_1$ | $K_{pb_1}$ | $NP_2$ | $K_{pb2}$ | 179.95 |
| 1 | -1 | $NP_1$ | f | $NP_2$ | f | 180.52 |
| 1 | -1 | N | $K_{pb_1}$ | N | $K_{pb2}$ | 180.56 |
| 1 | -1 | $NP_1$ | $P_1$ | $NP_2$ | $P_2$ | 180.53 |
| 1 | -1 | n | $P_1$ | n | $P_2$ | 179.49 |
| 1 | -1 | N | $\rho_{pb_1}$ | N | $\rho_{pb_2}$ | 180.55 |
| 1 | -1 | n | $P_1$ | n | $P_2$ | 180.36 |
| 1 | -1 | N | $P_1$ | N | $P_2$ | 180.57 |

**When,** $w_0 = 0 \; w_1 = 0 \; w_2 = 1$ also $L_i(i = 1,2,...,8) = 1$

$\alpha_1 = \alpha_2 = \beta_1 = \beta_2 = 1$ $PRE(t_p)$**=183.60**

## 5. *Double Sampling*

It is assumed that the population proportion $P_1$ for the first auxiliary attribute $\phi_1$ is unknown but the same is known for the second auxiliary attribute $\phi_2$. When $P_1$ is unknown, it is some times estimated from a preliminary large sample of size $n'$ on which only the attribute $\phi_1$ is measured. Then a second phase sample of size n ($n < n'$) is drawn and Y is observed.

Let $p'_j = \frac{1}{n}\sum_{i=1}^{n'} \phi_{ji}, (j = 1,2).$

The estimator's $t_1$, $t_2$, $t_3$ and $t_4$ in two-phase sampling take the following form

$$t_{d1} = \bar{y}\left(\frac{p_1'}{p_1}\right) \qquad (5.1)$$

$$t_{d2} = \bar{y}\left(\frac{P_2}{p_2^{'}}\right) \tag{5.2}$$

$$t_{d3} = \bar{y}\exp\left(\frac{p_1^{'} - p_1}{p_1^{'} + p_1}\right) \tag{5.3}$$

$$t_{d4} = \bar{y}\exp\left(\frac{p_2^{'} - P_2}{p_2^{'} + P_2}\right) \tag{5.4}$$

The bias and MSE expressions of the estimators $t_{d1}$, $t_{d2}$, $t_{d3}$ and $t_{d4}$ up to first order of approximation, are respectively given as

$$B(t_{d1}) = \bar{Y} f_3 C_{p_1}^2 \left[1 - k_{pb_1}\right] \tag{5.5}$$

$$B(t_{d2}) = \bar{Y} f_2 C_{p_2}^2 \left[1 - K_{pb_2}\right] \tag{5.6}$$

$$B(t_{d3}) = \bar{Y} f_3 \frac{C_{p_2}^2}{4}\left[1 - K_{pb_2}\right] \tag{5.7}$$

$$B(t_{d4}) = \bar{Y} f_3 \frac{C_{p2}^2}{4}\left[1 + K_{pb_2}\right] \tag{5.8}$$

$$MSE(t_{d1}) = \bar{Y}^2\left[f_1 C_y^2 + f_3 C_{P_1}^2 \left(1 - 2K_{pb_1}\right)\right] \tag{5.9}$$

$$MSE(t_{d2}) = \bar{Y}^2\left[f_1 C_y^2 + f_2 C_{p_2}^2 \left(1 - 2K_{kp_2}\right)\right] \tag{5.10}$$

$$MSE(t_{d3}) = \bar{Y}^2\left[f_1 C_y^2 + f_3 \frac{C_{p_1}^2}{4}\left(1 - 4K_{pb_1}\right)\right] \tag{5.11}$$

$$MSE(t_{d4}) = \bar{Y}^2\left[f_1 C_y^2 + f_3 \frac{C_{p_1}^2}{4}\left(1 + 4K_{pb_1}\right)\right] \tag{5.12}$$

where,

$$S_{\phi_J}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(\phi_{ji} - p_j\right)^2, \quad S_{\phi_j}^{'\,2} = \frac{1}{n^{'}-1}\sum_{i=1}^{n^{!}}\left(\phi_{ji} - p_j^{'}\right)^2,$$

$$f_2 = \frac{1}{n^{'}} - \frac{1}{N}, \quad f_3 = \frac{1}{n} - \frac{1}{n^{'}}.$$

The estimator's $t_5$ and $t_6$, in two-phase sampling, takes the following form

$$t_{d5} = \bar{y}\left(\frac{p_1'}{p_1}\right)^{m_1}\left(\frac{P_2}{p_2'}\right)^{m_2} \tag{5.13}$$

$$t_{d6} = \bar{y}\exp\left(\frac{p_1' - p_1}{p_1' + p_1}\right)^{n_1}\exp\left(\frac{p_2' - P_2}{p_2' + P_2}\right)^{n_2} \tag{5.14}$$

Where $m_1, m_2, n_1$ and $n_2$ are real constants.

The Bias and MSE expression's of the estimator's $t_{d5}$ and $t_{d6}$ up to the first order of approximation are, respectively, given by

$$B(t_{d5}) = \bar{Y}\left[f_3 C_{p_1}^2\left(\frac{m_1^2}{2} + \frac{m_1}{2} - m_1 K_{pb_1}\right) + f_2 C_{P_2}^2\left(\frac{m_2^2}{2} + \frac{m_2}{2} - m_2 k_{pb_2}\right)\right] \tag{5.15}$$

$$B(t_{d6}) = \bar{Y}\left[f_3\left(\frac{n_1^2}{8} + \frac{n_1}{8} - \frac{n_1}{2}K_{pb_1}\right)C_{p_1}^2 + f_2\left(\frac{n_2^2}{8} + \frac{n_2}{8} + \frac{n_2}{2}K_{pb_2}\right)\right] \tag{5.16}$$

$$MSE(t_{d5}) = \bar{Y}\left[f_1 C_y^2 + f_3 C_{p_1}^2\left(m_1^2 - 2m_1 K_{pb_1}\right) + f_2 C_{p_2 2}^2\left(m_2^2 - 2m_2 K_{pb_2}\right)\right] \tag{5.17}$$

$$MSE(t_{d6}) = \bar{Y}^2\left[f_1 C_y^2 + f_3\left(\frac{n_1^2}{4} - n_1 K_{pb_1}\right)C_{p_1}^2 + f_2\left(\frac{n_2^2}{4} + n_2 K_{pb_2}\right)C_{p_2}^2\right] \tag{5.18}$$

## 6. *Estimator $t_{pd}$ in Two-Phase Sampling*

Using linear combination of $t_{di}\,(i = 0,1,2)$, we define an estimator of the form

$$t_{pd} = \sum_{i=0}^{3} h_i t_{di} \in H \tag{6.1}$$

Such that, $\sum_{i=0}^{3} h_i = 1$ and $h_i \in R$ (6.2)

where,

$$t_0 = \bar{y},\ t_{d1} = \bar{y}\left[\frac{L_1 p'_1 + L_2}{L_1 p_1 + L_2}\right]^{m_1}\left[\frac{L_3 P_2 + L_4}{L_3 p'_2 + L_4}\right]^{m_2}$$

and $$t_{d2} = \exp\left[\frac{(L_5 p'_1 + L_6) - (Lp_1 + L_6)}{(L_1 p'_1 + L_2) + (L_5 p_1 + L_6)}\right]^{n_1}\exp\left[\frac{(L_7 p'_2 + L_6) - (L_7 P_2 + L_8)}{(L_7 p'_2 + L_2) + (L_7 P_2 + L_8)}\right]^{n_2}$$

where $h_i\,(i = 0,1,2)$ denotes the constants used for reducing the bias in the class of estimators, H denotes the set of those estimators that can be constructed from $t_{di}\,(i = 0,1,2)$ and R denotes the set of real numbers (for detail see Singh et. al. (2008)). Also, $L_i\left(i=1,2,...,8\right)$ are either real numbers or the functions of the known parameters of the auxiliary attributes.

Expressing $t_{pd}$ in terms of e's, we have

$$t_p = \overline{Y}(1+e_0)\left[h_0 + h_1(1+\varphi_1 e'_1)^{m_1}(1+\varphi_1 e_1)^{-m_1}(1+\varphi_2 e'_2)^{-m_2}\right.$$
$$\left. + h_2 \exp\left(\theta_1[e'_1 - e_1][1+\theta_1(e'_1-e_1)]^{-1}\right)^{n_1} \exp\left(\theta_2 e'_2[1+\theta_2 e'_2]\right)^{n_2}\right] \quad (6.3)$$

After expanding, subtracting $\overline{Y}$ from both sides of the equation (6.3) and neglecting the terms having power greater than two, we have

$$\left(t_{pd} - \overline{Y}\right) = \overline{Y}\left[e_0 + h_1\left(m_1\varphi_1 e'_1 - m_1\varphi_1 e_1 - m_2\varphi_2 e'_2\right) + h_2\left(n_1\theta_1 e'_1 - n_1\theta_1 e_1 + n_2\theta_2 e'_2\right)\right] \quad (6.4)$$

Squaring both sides of (6.4) and then taking expectations, we get MSE of the estimator $t_p$ up to the first order of approximation, as

$$MSE\left(t_{pd}\right) = \overline{Y}^2\left[h_1^2 R_1 + h_2^2 R_2 + 2h_1 h_2 R_3 + 2h_1 R_4 + 2h_2 R_5\right] \quad (6.5)$$

where,

$$\left.\begin{aligned} h_1 &= \frac{R_2 R_4 - R_3 R_5}{R_1 R_2 - R_3^2} \\ h_2 &= \frac{R_1 R_5 - R_3 R_4}{R_1 R_2 - R_3^2} \end{aligned}\right\} \quad (6.6)$$

and

$$\left.\begin{aligned} R_1 &= \varphi_1^2 m_1^2 f_3 C_{p_1}^2 + \varphi_2^2 m_2^2 f_2 C_{p_2}^2 \\ R_2 &= \theta_1^2 n_1^2 f_3 C_{p_1}^2 + \theta_2^2 n_2^2 f_2 C_{p_2}^2 \\ R_3 &= m_2 n_2 f_2 \varphi_2 \theta_2 C_{p_2}^2 - n_1 m_1 \varphi_1 \theta_1 f_2 k_\varphi C_{p_1}^2 \\ R_4 &= -m_1\varphi_1 f_3 k_{pb_1} C_{p_1}^2 - m_2\varphi_2 f_2 k_{pb_2} C_{p_2}^2 \\ R_5 &= -n_1\theta_1 f_3 k_{pb_1} C_{p_1}^2 + n_2\theta_2 f_2 k_{pb_2} C_{p_2}^2 \end{aligned}\right\} \quad (6.7)$$

Data: (Source: Singh and Chaudhary (1986), p. 177).

The population consists of 34 wheat farms in 34 villages in certain region of India. The variables are defined as:

y = area under wheat crop (in acres) during 1974.

$p_1$ = proportion of farms under wheat crop which have more than 500 acres land during 1971. and

$p_2$ = proportion of farms under wheat crop which have more than 100 acres land during 1973.

For this data, we have

N=34, $\bar{Y}$ =199.4, $P_1$=0.6765, $P_2$=0.7353, $S_y^2$=22564.6, $S_{\phi_1}^2$ =0.225490, $S_{\phi_2}^2$ =0.200535,

$\rho_{pb_1}$ =0599, $\rho_{pb_2}$ =0.559, $\rho_\phi$=0.725.

**Table 6.1: PRE of different estimators of $\bar{Y}$ with respect to $\bar{y}$**

| **CHOICE OF SCALERS, when** $h_0 = 0\ h_1 = 1\ h_2 = 0$ | | | | | | |
|---|---|---|---|---|---|---|
| $m_1$ | $m_2$ | $L_1$ | $L_2$ | $L_3$ | $L_4$ | PRE'S |
| 0 | 1 | | | 1 | 0 | 108.16 |
| 1 | 0 | 1 | 0 | | | 121.59 |
| 1 | 1 | 1 | 1 | 1 | 1 | 142.19 |
| 1 | 1 | 1 | 0 | 1 | 0 | 133.40 |
| 1 | 1 | $C_{p_1}$ | $\rho_{pb_1}$ | $C_{p_2}$ | $\rho_{pb_2}$ | 144.78 |
| 1 | 1 | $NP_1$ | $K_{pb_1}$ | $NP_2$ | $K_{pb2}$ | 136.90 |
| 1 | 1 | $NP_1$ | f | $NP_2$ | f | 133.30 |
| 1 | 1 | N | $K_{pb_1}$ | N | $K_{pb2}$ | 135.73 |
| 1 | 1 | $NP_1$ | $P_1$ | $NP_2$ | $P_2$ | 137.09 |
| 1 | 1 | n | $P_1$ | n | $P_2$ | 138.23 |
| 1 | 1 | N | $\rho_{pb_1}$ | N | $\rho_{pb_2}$ | 135.49 |
| 1 | 1 | n | $P_1$ | n | $P_2$ | 138.97 |
| 1 | 1 | N | $P_1$ | N | $P_2$ | 135.86 |
| **When,** $h_0 = 0\ h_1 = 0\ h_2 = 1$ | | | | | | |
| $n_1$ | $n_2$ | $L_5$ | $L_6$ | $L_7$ | $L_8$ | PRE'S |
| 1 | 0 | 1 | 0 | 1 | 0 | 130.89 |
| 0 | -1 | 1 | 0 | 1 | 0 | 108.93 |
| 1 | -1 | 1 | 0 | 1 | 0 | 146.63 |
| 1 | -1 | 1 | 1 | 1 | 1 | 121.68 |
| 1 | -1 | 1 | 1 | 1 | 0 | 127.24 |
| 1 | -1 | $C_{p_1}$ | $\rho_{pb_1}$ | $C_{p_2}$ | $\rho_{pb_2}$ | 123.43 |
| 1 | -1 | $NP_1$ | $K_{pb_1}$ | $NP_2$ | $K_{pb2}$ | 145.49 |
| 1 | -1 | $NP_1$ | f | $NP_2$ | f | 146.57 |
| 1 | -1 | N | $K_{pb_1}$ | N | $K_{pb2}$ | 145.84 |
| 1 | -1 | $NP_1$ | $P_1$ | $NP_2$ | $P_2$ | 145.43 |
| 1 | -1 | n | $P_1$ | n | $P_2$ | 145.03 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 1 | -1 | N | $\rho_{pb_1}$ | N | $\rho_{pb_2}$ | 145.92 |
| 1 | -1 | n | $P_1$ | n | $P_2$ | 144.85 |
| 1 | -1 | N | $P_1$ | N | $P_2$ | 145.80 |
| **When,** $h_0 = 0\ h_1 = 0\ h_2 = 1$ also $L_i(i = 1,2,...,8) = 1$ $m_1 = m_2 = n_1 = n_2 = 1$ | | | | | | $PRE(t_{pd})$=**154.28** |

## *7. Conclusion*

In this paper, we have suggested a class of estimators in single and two-phase sampling by using point bi serial correlation and phi correlation coefficient. From Table 4.1 and Table 6.1, we observe that the proposed estimator $t_p$ and $t_{pd}$ performs better than other estimators considered in this paper.

## *References*


1. Abd-Elfattah, A.M. El-Sherpieny, E.A. Mohamed, S.M. Abdou, O. F., 2010, Improvement in estimating the population mean in simple random sampling using information on auxiliary attribute. Appl. Mathe. and Compt. doi:10.1016/j.amc.2009.12.041

2. Government of Pakistan, 2004, Crops Area Production by Districts (Ministry of Food, Agriculture and Livestock Division, Economic Wing, Pakistan).

3. Gujarati, D. N. and Sangeetha, 2007, Basic econometrics. Tata McGraw – Hill.

4. Jhajj, H.S., Sharma, M.K. and Grover, L.K., 2006 , A family of estimators of population mean using information on auxiliary attribute. Pak. Journ. of Stat., 22(1), 43-50.

5. Malik, S. And Singh, R. ,2012, A Family Of Estimators Of Population Mean Using Information On Point Bi-Serial And Phi-Correlation Coefficient. Intern. Jour. Stat. And Econ. (accepted).

6. Naik,V.D and Gupta, P.C., 1996, A note on estimation of mean with known population proportion of an auxiliary character. Jour. Ind. Soc. Agri. Stat., 48(2), 151-158.

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

8. Singh, D. and Chaudhary, F. S., 1986, Theory and Analysis of Sample Survey Designs (John Wiley and Sons, NewYork).

9. Singh, R., Cauhan, P., Sawan, N. and Smarandache, F., 2007, Auxiliary information and a priori values in construction of improved estimators. Renaissance High press.

10. 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(1) 47–53.

11. Singh, R., Kumar, M. and Smarandache, F., 2010, Ratio estimators in simple random sampling when study variable is an attribute. WASJ 11(5): 586-589.

12. Yule, G. U., 1912, On the methods of measuring association between two attributes. Jour. of The Royal Soc. 75, 579-642.