SAMPLING STRATEGIES FOR FINITE POPULATION USING AUXILIARY INFORMATION

Florentin Smarandache
University of New Mexico, smarand@unm.edu

Rajesh Singh

Follow this and additional works at: https://digitalrepository.unm.edu/math_fsp

Part of the Agricultural and Resource Economics Commons, Counseling Commons, Environmental Studies Commons, Income Distribution Commons, Mathematics Commons, and the Public Affairs, Public Policy and Public Administration Commons

Recommended Citation
Rajesh Singh, Florentin Smarandache. SAMPLING STRATEGIES FOR FINITE POPULATION USING AUXILIARY INFORMATION. Ohio: The Educational Publisher, 2015

This Book is brought to you for free and open access by the Mathematics at UNM Digital Repository. It has been accepted for inclusion in Faculty and Staff Publications by an authorized administrator of UNM Digital Repository. For more information, please contact amywinter@unm.edu, lsloane@salud.unm.edu, sarahrk@unm.edu.
Rajesh Singh, Florentin Smarandache (editors)

SAMPLING STRATEGIES FOR FINITE POPULATION
USING AUXILIARY INFORMATION

Papers by Sachin Malik, Rajesh Singh, Florentin Smarandache,
The Educational Publisher
Zip Publishing
1313 Chesapeake Ave.
Columbus, Ohio 43212, USA
Email: info@edupublisher.com


© The Authors, The Editors, The Publisher, 2015.
Sampling Strategies for Finite Population Using Auxiliary Information

Rajesh Singh
Department of Statistics, BHU, Varanasi (U.P.), India
Editor

Florentin Smarandache
Chair of Department of Mathematics, University of New Mexico, Gallup, USA
Editor

Sampling Strategies
For Finite Population
Using Auxiliary Information

The Educational Publisher
Columbus, 2015
# Contents

Foreword ................................................................................................................................... 7

A Generalized Family Of Estimators For Estimating Population Mean Using Two Auxiliary Attributes ................................................................. 9

Abstract ................................................................................................................................ 9
Keywords ............................................................................................................................... 9
1. Introduction ....................................................................................................................... 9
2. Some Estimators in Literature .......................................................................................... 10
3. The Suggested Class of Estimators ................................................................................. 12
4. Empirical Study ................................................................................................................ 14
5. Double Sampling ............................................................................................................. 15
6. Estimator $t_{pd}$ in Two-Phase Sampling ......................................................................... 17
7. Conclusion ....................................................................................................................... 20
References ............................................................................................................................ 20

A General Procedure of Estimating Population Mean Using Information on Auxiliary Attribute .................................................................................................................................. 21

Abstract ................................................................................................................................ 21
Keywords ............................................................................................................................. 21
1. Introduction ...................................................................................................................... 21
2. Proposed Estimator .......................................................................................................... 22
3. Members of the family of estimator of $t$ and their Biases and MSE ............................ 25
4. Empirical study ................................................................................................................ 28
Conclusion ............................................................................................................................ 29
References ............................................................................................................................ 29

Estimation of Ratio and Product of Two Population Means Using Auxiliary Characters in the Presence of Non Response .......................................................................................... 31

Abstract ................................................................................................................................ 31
Keywords ............................................................................................................................. 31
Introduction .......................................................................................................................... 31
Estimation of Ratio and product of two population means .................................................. 31
    Case 1. The Case of Complete Response: ........................................................................ 31
    Case 2. Incomplete Response in the Sample due to Non-response: ................................. 34
References ............................................................................................................................ 36

On The Use of Coefficient of Variation and $\beta_1, \beta_2$ in Estimating Mean of a Finite Population ........................................................................................................................................ 39

Abstract ................................................................................................................................ 39
Keywords ............................................................................................................................. 39
Introduction .......................................................................................................................... 39
Estimators and their Mean Square Error .............................................................................. 39
References ............................................................................................................................ 43
Foreword

The present book aims to present some improved estimators using auxiliary and attribute information in case of simple random sampling and stratified random sampling and in some cases when non-response is present.

This volume is a collection of five papers, written by seven co-authors (listed in the order of the papers): Sachin Malik, Rajesh Singh, Florentin Smarandache, B. B. Khare, P. S. Jha, Usha Srivastava and Habib Ur. Rehman.

The first and the second papers deal with the problem of estimating the finite population mean when some information on two auxiliary attributes are available. In the third paper, problems related to estimation of ratio and product of two population mean using auxiliary characters with special reference to non-response are discussed.

In the fourth paper, the use of coefficient of variation and shape parameters in each stratum, the problem of estimation of population mean has been considered. In the fifth paper, a study of improved chain ratio-cum-regression type estimator for population mean in the presence of non-response for fixed cost and specified precision has been made.

The authors hope that the book will be helpful for the researchers and students that are working in the field of sampling techniques.
A Generalized Family Of Estimators For Estimating Population Mean Using Two Auxiliary Attributes

Sachin Malik, Rajesh Singh and Florentin Smarandache

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

2Chair 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 biserial 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 biserial 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 $\bar{Y}$ 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:

\[
t_1 = \frac{-\bar{y}P_1}{p_1}
\]  
(2.1)

\[
t_2 = \frac{-\bar{y}P_2}{p_2}
\]  
(2.2)

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

\[
t_4 = \bar{y}\exp\left(\frac{p_2 - P_2}{p_2 + P_2}\right)
\]  
(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

\[
B(t_i) = \bar{y}f_1 C_p^2 \left[1 - K_{p_1}\right]
\]  
(2.5)

\[
B(t_2) = \bar{y}f_1 K_{p_2} C_{p_2}
\]  
(2.6)

\[
B(t_3) = \bar{y}f_1 \frac{C_p^2}{2} \left[\frac{1}{4} - K_{p_2}\right]
\]  
(2.7)

\[
B(t_4) = \bar{y}f_1 \frac{C_p^2}{2} \left[\frac{1}{4} + K_{p_2}\right]
\]  
(2.8)

\[
\text{MSE}(t_1) = \bar{y}^2 f_1 \left[C_y^2 + C_p^2 \left(1 - 2K_{p_1}\right)\right]
\]  
(2.9)

\[
\text{MSE}(t_2) = \bar{y}^2 f_1 \left[C_y^2 + C_p^2 \left(1 + 2K_{p_2}\right)\right]
\]  
(2.10)
Sampling Strategies for Finite Population Using Auxiliary Information

\[ \text{MSE}(t_3) = \bar{Y}^2 f_i \left[ C_y^2 + C_{p_1}^2 \left( \frac{1}{4} - K_{pb_1} \right) \right] \]  
(2.11)

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

where, \( f_i = \frac{1}{n} - \frac{1}{N} \), \( S_{\phi j} = \frac{1}{N-1} \sum_{i=1}^{N} (\phi_{ji} - P_j)^2 \), \( S_{yj} = \frac{1}{N-1} \sum_{i=1}^{N} (y_i - \bar{Y})(\phi_{ji} - P_j) \), 
\[ \rho_{pb_j} = \frac{S_{yj}}{S_{\phi j}} \] \( C_j = \frac{S_y}{P_j} \), \( C_{p_j} = \frac{S_{\phi j}}{P_j} \); \( j = 1, 2 \),
\[ K_{pb_1} = \rho_{pb_1} \frac{C_y}{C_{p_1}} \]
\[ K_{pb_2} = \rho_{pb_2} \frac{C_y}{C_{p_2}} \]

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

and \( \rho_{\phi} = \frac{S_{\phi:2}}{S_{\phi_1} S_{\phi_2}} \).

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

\[ t_5 = \bar{y} \left( \frac{P_1}{P_1} \right)^{a_1} \left( \frac{P_2}{P_2} \right)^{a_2} \]  
(2.13)

\[ t_6 = \bar{y} \exp \left( \frac{P_1 - P_1}{P_1 + P_1} \right)^{b_1} \exp \left( \frac{P_2 - P_2}{P_2 + P_2} \right)^{b_2} \]  
(2.14)

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

The Bias and MSE expressions of the estimator’s \( t_5 \) and \( t_6 \) up to the first order of approximation are, respectively, given by

\[ B(t_5) = \bar{Y} f_i \left[ C_y^2 \left( \frac{\alpha_1^2}{2} + \frac{\alpha_1}{2} - \alpha_1 k_{pb_1} \right) + C_{p_1}^2 \left( \frac{\alpha_2^2}{2} + \frac{\alpha_2}{2} - \alpha_2 k_{pb_2} + \alpha_1 \alpha_2 k_{\phi} \right) \right] \]  
(2.15)
3. The Suggested Class of Estimators

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

\[
t_p = \sum_{i=0}^{3} w_i t_i \in H
\]

(3.1)

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

(3.2)

Where,

\[
t_0 = \overline{y} , \quad t_1 = \overline{y} \left[ \frac{L_1 p_1 + L_2}{L_1 p_1 + L_2} \right]^{\alpha_1} \left[ \frac{L_3 p_2 + L_4}{L_3 p_2 + L_4} \right]^{\alpha_2}
\]

and

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

where \( w_i (i = 0,1,2) \) 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 \( t_i (i = 0,1,2) \) and \( R \) denotes the set of real numbers (for detail see Singh et. al (2008)). Also, \( L_i (i = 1,2,\ldots,8) \) 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

\[
t_p = \overline{Y} (1 + \epsilon_0) \left[ w_0 + w_1 (1 + \varphi_1 \epsilon_1)^{-\alpha_1} (1 + \varphi_2 \epsilon_2)^{-\alpha_2} \right] + w_2 \exp \left( -\theta_1 \epsilon_1 \left[ 1 + \theta_1 \epsilon_1 \right]^{-1} \right)^{\beta_1} \exp \left( -\theta_2 \epsilon_2 \left[ 1 + \theta_2 \epsilon_2 \right]^{-1} \right)^{\beta_2}
\]

(3.3)

where,
\[ \varphi_1 = \frac{L_1 P_1}{L_1 P_1 + L_2} \]
\[ \varphi_2 = \frac{L_3 P_2}{L_3 P_1 + L_4} \]
\[ \theta_1 = \frac{L_4 P_1}{2[L_3 P_2 + L_6]} \]
\[ \theta_2 = \frac{L_5 P_2}{2[L_3 P_2 + L_8]} \]

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

\[ (t_p - \overline{Y}) = \sqrt{\overline{e_0}} - \sqrt{w_1(\alpha_i \varphi_{i1} e_i + \alpha_2 \varphi_{i2} e_2)} - \sqrt{w_2(\beta_1 \varphi_{i1} e_i - \beta_2 \varphi_{i2} e_2)} ] \]

(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

\[ \text{MSE}(t_p) = \overline{Y}^2 f[w_1^2 T_1 + w_2^2 T_2 + 2w_1 w_2 T_3 - 2w_1 T_4 - 2w_2 T_5] \]

(3.5)

where,

\[ w_1 = \frac{L_2 L_4 - L_3 L_5}{L_1 L_2 - L_3^2} \]
\[ w_2 = \frac{L_1 L_4 - L_3 L_4}{L_1 L_2 - L_3^2} \]

(3.6)

and

\[ L_1 = \varphi_1^2 \alpha_1^2 C_{p_1} + \varphi_2^2 \alpha_2^2 C_{p_2} + 2\alpha_1 \alpha_2 \varphi_{i1} \varphi_{i2} k_\varphi C_{p_2}^2 \]
\[ L_2 = \theta_1^2 \beta_1^2 C_{p_1} + \theta_2^2 \beta_2^2 C_{p_2} - 2\beta_1 \beta_2 \varphi_{i1} \varphi_{i2} k_\varphi C_{p_2} \]
\[ L_3 = \alpha_1 \beta_1 \varphi_{i1} C_{p_1} - \alpha_2 \beta_1 \varphi_{i2} C_{p_2} + \alpha_2 \beta_1 \varphi_{i1} \varphi_{i2} k_\varphi C_{p_2}^2 - \alpha_1 \varphi_{i1} \beta_2 \varphi_{i2} k_\varphi C_{p_2} \]
\[ L_4 = \alpha_1 \varphi_{i1} k_{p_0} C_{p_1}^2 + \alpha_2 \varphi_{i2} k_{p_0} C_{p_2}^2 \]
\[ L_5 = \beta_1 \varphi_{i1} k_{p_0} C_{p_1}^2 - \beta_2 \varphi_{i2} k_{p_0} C_{p_2}^2 \]

(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\(^{th}\) 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

\( N=73, \bar{Y}=61.3, P_1=0.4247, P_2=0.3425, S_y^2=12371.4, S_{\theta_1}^2=0.225490, S_{\theta_2}^2=0.228311,\)

\( \rho_{p_1}=0.621, \rho_{p_2}=0.673, \rho_{\theta}=0.889.\)

<table>
<thead>
<tr>
<th>CHOICE OF SCALERS, when ( w_0 = 0 ) ( w_1 = 1 ) ( w_2 = 0 )</th>
<th>PRE’S</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha_1 )</td>
<td>( \alpha_2 )</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>-1</td>
<td>1</td>
</tr>
</tbody>
</table>

When, \( w_0 = 0 \) \( w_1 = 0 \) \( w_2 = 1 \)
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 sometimes 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

$$t_{di} = \frac{y}{y \left( \frac{p_i'}{p_i} \right)}$$

(5.1)
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

$$B(t_{d1}) = \bar{Y}f_1C_{p_1}^2 \left[ 1 - k_{pb_1} \right]$$  \hspace{1cm} (5.5)

$$B(t_{d2}) = \bar{Y}f_2C_{p_2}^2 \left[ 1 - k_{pb_2} \right]$$  \hspace{1cm} (5.6)

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

$$B(t_{d4}) = \bar{Y}f_3 \frac{C_{p_2}^2}{4} \left[ 1 + K_{pb_2} \right]$$  \hspace{1cm} (5.8)

$$\text{MSE}(t_{d1}) = \bar{Y}^2 \left[ f_1C_y^2 + f_3C_{p_1}^2 \left( 1 - 2K_{pb_1} \right) \right]$$  \hspace{1cm} (5.9)

$$\text{MSE}(t_{d2}) = \bar{Y}^2 \left[ f_1C_y^2 + f_2C_{p_2}^2 \left( 1 - 2K_{pb_2} \right) \right]$$  \hspace{1cm} (5.10)

$$\text{MSE}(t_{d3}) = \bar{Y}^2 \left[ f_1C_y^2 + f_3 \frac{C_{p_2}^2}{4} \left( 1 - 4K_{pb_1} \right) \right]$$  \hspace{1cm} (5.11)

$$\text{MSE}(t_{d4}) = \bar{Y}^2 \left[ f_1C_y^2 + f_3 \frac{C_{p_1}^2}{4} \left( 1 + 4K_{pb_1} \right) \right]$$  \hspace{1cm} (5.12)

where,

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

$$f_2 = \frac{1}{n} - \frac{1}{N}, \quad f_3 = \frac{1}{n} - \frac{1}{n}.$$
The estimator’s $t_5$ and $t_6$, in two-phase sampling, takes the following form

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

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

(5.13)

(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

$$B(t_{d5}) = \bar{Y} \left[ f_1 C_{p_1} \left( \frac{m_1^2}{2} + \frac{m_1}{2} - m_1 K_{p_1} \right) + f_2 C_{p_1} \left( \frac{m_2^2}{2} + \frac{m_2}{2} - m_2 K_{p_2} \right) \right]$$

(5.15)

$$B(t_{d6}) = \bar{Y} \left[ f_3 \left( \frac{n_1^2}{8} + \frac{n_1}{2} - n_1 K_{p_1} \right) C_{p_1} + f_2 \left( \frac{n_2^2}{8} + \frac{n_2}{2} + n_2 K_{p_2} \right) \right]$$

(5.16)

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

(5.17)

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

(5.18)

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

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

$$t_{pd} = \sum_{i=0}^{3} h_i t_{d_i} \in H$$

(6.1)

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

(6.2)

where,

$$t_0 = \bar{y}, \quad t_{d1} = \bar{y} \left[ \frac{L_1 p_1' + L_2}{L_1 p_1 + L_2} \right]^{m_1} \left[ \frac{L_3 p_2 + L_4}{L_3 p_2' + L_4} \right]^{m_2}$$

$$t_{d2} = \exp \left[ \left( \frac{L_1 p_1' + L_4}{L_1 p_1 + L_4} \right) - \left( \frac{L_1 p_1' + L_4}{L_1 p_1 + L_4} \right) \right]^{h_1} \exp \left[ \left( \frac{L_2 p_2' + L_6}{L_2 p_2 + L_6} \right) - \left( \frac{L_2 p_2' + L_6}{L_2 p_2 + L_6} \right) \right]^{h_2}$$

(6.3)

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 $t_{d_i} (i = 0,1,2)$ and $R$
denotes the set of real numbers (for detail see Singh et. al. (2008)). Also, \( L_i (i = 1, 2, \ldots, 8) \) 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

\[
t_p = \overline{Y} (1 + e_0) \left[ h_0 + h_1 (1 + \phi_1 e'_1)^m_1 (1 + \phi_1 e'_1)^2 (1 + \phi_2 e'_2)^m_2 \right] + h_2 \exp \left( \theta_1 (e'_1 - e_1 (1 + \theta_1 (e'_1 - e_1))^n_1 \right) \exp \left( \theta_2 e'_2 \right) \left[ 1 + \theta_2 (e'_2)^n_2 \right] (6.3)
\]

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

\[
\left( t_{pd} - \overline{Y} \right) = \overline{Y} \left[ e_0 + h_1 (m_1 \phi_1 e'_1 - m_1 \phi_1 e_1 - m_2 \phi_2 e'_2) + h_2 (n_1 \theta_1 e'_1 - n_1 \theta_1 e_1 + n_2 \theta_2 e'_2) \right] \tag{6.4}
\]

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

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

where,

\[
\begin{align*}
  h_1 &= \frac{R_2 R_4 - R_3 R_5}{R_1 R_2 - R_3^2} \\
  h_2 &= \frac{R_1 R_4 - R_1 R_5}{R_1 R_2 - R_3^2}
\end{align*} \tag{6.6}
\]

and

\[
\begin{align*}
  R_1 &= \phi_1^2 m_1^2 f_1 C_{p_1}^2 + \phi_2^2 m_2^2 f_2 C_{p_2}^2 \\
  R_2 &= \theta_1^2 n_1^2 f_3 C_{p_1}^2 + \theta_2^2 n_2^2 f_2 C_{p_2}^2 \\
  R_3 &= m_2 n_2 f_2 \phi_2 \theta_2 C_{p_2}^2 - n_1 m_1 \phi_1 \theta_1 f_1 k_1 C_{p_1}^2 \\
  R_4 &= -m_1 \phi_1 f_3 k_{p_1} C_{p_1}^2 - m_2 \phi_2 f_2 k_{p_2} C_{p_2}^2 \\
  R_5 &= -n_1 \theta_1 f_3 k_{p_1} C_{p_1}^2 + n_2 \theta_2 f_2 k_{p_2} C_{p_2}^2
\end{align*} \tag{6.7}
\]


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.

\( p_i = \) proportion of farms under wheat crop which have more than 500 acres land during 1971.
p_2 = proportion of farms under wheat crop which have more than 100 acres land during 1973.

For this data, we have

N=34, \( \bar{Y} = 199.4 \), \( P_1 = 0.6765 \), \( P_2 = 0.7353 \), \( S^2_y = 22564.6 \), \( S^2_{\phi_1} = 0.225490 \), \( S^2_{\phi_2} = 0.200535 \),

\( \rho_{ph_1} = 0.599 \), \( \rho_{ph_2} = 0.559 \), \( \phi = 0.725 \).

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

| CHOICE OF SCALERS, when \( h_0 = 0 \) \( h_1 = 1 \) \( h_2 = 0 \) | \( h_0 = 0 \) \( h_1 = 0 \) \( h_2 = 1 \) |
|---|---|---|---|---|---|---|
| \( m_1 \) | \( m_2 \) | \( L_1 \) | \( L_2 \) | \( L_3 \) | \( L_4 \) | PRE’S |
| 0 | 1 | 1 | 0 | 0 | 0 | 108.16 |
| 1 | 0 | 1 | 0 | 1 | 0 | 121.59 |
| 1 | 1 | 1 | 1 | 1 | 1 | 142.19 |
| 1 | 1 | 0 | 1 | 0 | 1 | 133.40 |
| 1 | 1 | \( C_{p_1} \) | \( \rho_{ph_1} \) | \( C_{p_2} \) | \( \rho_{ph_2} \) | 144.78 |
| 1 | 1 | \( NP_1 \) | \( K_{ph_1} \) | \( NP_2 \) | \( K_{ph_2} \) | 136.90 |
| 1 | 1 | \( NP_1 \) | \( n \) | \( NP_2 \) | \( n \) | 133.30 |
| 1 | 1 | \( N \) | \( K_{ph_1} \) | \( N \) | \( K_{ph_2} \) | 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 | \( N \) | \( \rho_{ph_1} \) | \( N \) | \( \rho_{ph_2} \) | 135.49 |
| 1 | 1 | \( n \) | \( P_1 \) | \( n \) | \( P_2 \) | 138.97 |
| 1 | 1 | \( N \) | \( P_1 \) | \( N \) | \( P_2 \) | 135.86 |

When, \( h_0 = 0 \) \( h_1 = 0 \) \( h_2 = 1 \)

<table>
<thead>
<tr>
<th>( n_1 )</th>
<th>( n_2 )</th>
<th>( L_5 )</th>
<th>( L_6 )</th>
<th>( L_7 )</th>
<th>( L_8 )</th>
<th>PRE’S</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>130.89</td>
</tr>
<tr>
<td>0</td>
<td>-1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>108.93</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>146.63</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>121.68</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>127.24</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>( C_{p_1} )</td>
<td>( \rho_{ph_1} )</td>
<td>( C_{p_2} )</td>
<td>( \rho_{ph_2} )</td>
<td>123.43</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>( NP_1 )</td>
<td>( K_{ph_1} )</td>
<td>( NP_2 )</td>
<td>( K_{ph_2} )</td>
<td>145.49</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>( NP_1 )</td>
<td>( f )</td>
<td>( NP_2 )</td>
<td>( f )</td>
<td>146.57</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>( N )</td>
<td>( K_{ph_1} )</td>
<td>( N )</td>
<td>( K_{ph_2} )</td>
<td>145.84</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>( NP_1 )</td>
<td>( P_1 )</td>
<td>( NP_2 )</td>
<td>( P_2 )</td>
<td>145.43</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>( n )</td>
<td>( P_1 )</td>
<td>( n )</td>
<td>( P_2 )</td>
<td>145.03</td>
</tr>
</tbody>
</table>
When, \[ h_0 = 0 \ h_1 = 0 \ h_2 = 1 \] also \[ L_i(i = 1, 2, ..., 8) = 1 \]
\[ m_1 = m_2 = n_1 = n_2 = 1 \]

\[ \text{PRE}(t_{pd}) = 154.28 \]

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

A General Procedure of Estimating Population Mean Using
Information on Auxiliary Attribute

1Sachin Malik, 11Rajesh Singh and 2Florentin Smarandache

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

2Chair 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 auxiliary attribute is available. It is shown that the proposed estimator is more
efficient than the usual mean estimator and other existing estimators. The results have been
illustrated numerically by taking empirical population considered in the literature.

Keywords Simple random sampling, auxiliary attribute, point bi-serial correlation, ratio
estimator, efficiency.

1. Introduction
The use of auxiliary information can increase the precision of an estimator when
study variable y is highly correlated with auxiliary variable x. There are many situations
when auxiliary information is available in the form of attributes, e.g. sex and height of the
persons, amount of milk produced and a particular breed of cow, amount of yield of wheat
crop and a particular variety of wheat (see Jhajj et. al. (2006)).

Consider a sample of size n drawn by simple random sampling without replacement
(SRSWOR) from a population of size N. Let \( y_i \) and \( \phi_i \) denote the observations on variable y
and \( \phi \) respectively for \( i^{th} \) unit ( \( i = 1, 2, \ldots, N \)).

Let \( \phi_i = 1 \); if the \( i^{th} \) unit of the population possesses attribute \( \phi = 0 \); otherwise.

Let \( A = \sum_{i=1}^{N} \phi_i \) and \( a = \sum_{i=1}^{n} \phi_i \), denote the total number of units in the population and sample
respectively possessing attribute \( \phi \). Let \( P = A/N \) and \( p = a/n \) denote the proportion of units in
the population and sample respectively possessing attribute \( \phi \). Naik and Gupta (1996)
introduced a ratio estimator \( t_{NG} \) when the study variable and the auxiliary attribute are
positively correlated. The estimator \( t_{NG} \) is given by
$t_{NG} = \frac{y - p}{p}$  \hspace{1cm} (1.1)

with MSE

$$\text{MSE}(t_{NG}) = f_1 \left( S_y^2 + R S_\phi^2 - 2RS_{y\phi} \right)$$  \hspace{1cm} (1.2)

where $f_1 = \frac{N-n}{Nn}$, $R = \frac{\bar{Y}}{p}$, $S_y^2 = \frac{1}{N-n} \sum_{i=1}^{N} (y_i - \bar{Y})^2$

$$S_\phi^2 = \frac{1}{N-1} \sum_{i=1}^{N} (\phi_i - p) \quad S_{y\phi} = \frac{1}{N-1} \sum_{i=1}^{N} (\phi_i - p)(y_i - \bar{Y})$$

(for details see Singh et al. (2008))

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) and Abd-Elfattah et al. (2010) have considered the problem of estimating population mean $\bar{Y}$ taking into consideration the point biserial correlation coefficient between auxiliary attribute and study variable.

The objective of this article is to suggest a generalised class of estimators for population mean $\bar{Y}$ and analyse its properties. A numerical illustration is given in support of the present study.

2. Proposed Estimator

Let $\phi_i^* = \phi_i + mA$, $m$ being a suitably chosen scalar, that takes values 0 and 1. Then

$$q = p + mA = p + NmP \quad \text{and} \quad Q = (Nm + 1)P,$$

where $q = \frac{b}{n}, Q = \frac{B}{N}, B = \sum_{i=1}^{N} \phi_i$ and $b = \sum_{i=1}^{N} \phi_i$.

Motivated by Bedi (1996), we define a family of estimators for population mean $\bar{Y}$ as

$$t = \left[ w_1 \bar{Y} + w_2 b(P - p) \left( \frac{q}{Q} \right) ^\alpha \right]$$  \hspace{1cm} (2.1)

where $w_1, w_2$ and $\alpha$ are suitably chosen scalars.

To obtain the Bias and MSE of the estimator $t$, we write
\[ \bar{y} = \bar{Y}(1 + e_0), \quad p = P(1 + e_1), \quad s_{\phi}^2 = S_{\phi}^2(1 + e_2), \]

\[ s_{\phi y} = S_{\phi y}(1 + e_3), \quad b = \beta(1 + e_3)(1 + e_2)^{-1} \]

such that

\[ E(e_i) = 0, \quad i=0,1,2,3 \]

and

\[ E(e_0^2) = \left( \frac{1}{n} - \frac{1}{N} \right) C_y^2, \]

\[ E(e_1^2) = \left( \frac{1}{n} - \frac{1}{N} \right) C_p^2, \]

\[ E(e_0e_1) = \left( \frac{1}{n} - \frac{1}{N} \right) \rho_{pb} C_y C_p, \]

\[ E(e_1^2) = \left( \frac{1}{n} - \frac{1}{N} \right) C_p \lambda_{03}, \]

\[ E(e_1e_3) = \left( \frac{1}{n} - \frac{1}{N} \right) C_p \frac{\lambda_{12}}{\rho_{pb}}, \]

Expressing (2.1) in terms of \( e \)'s, we have

\[ t = \bar{Y} \left[ w_1(1 + e_0) - w_2 \frac{\beta}{R} e_1(1 + e_3)(1 + e_2)^{-1} \left( 1 + \frac{e_1}{Nm + 1} \right)^\alpha \right] \]

(2.2)

We assume that \( |e_i| < 1 \) and \( \frac{e_1}{Na + 1} < 1 \), so that \( (1 + e_3)^{-1} \) and \( \left( 1 + \frac{e_1}{Nm + 1} \right)^\alpha \) are expandable.

Expanding the right hand side of (2.2) and retaining terms up to second powers of \( e \)'s, we have

\[ t - \bar{Y} = \bar{Y} \left[ w_1 \left\{ 1 + e_0 + \frac{\alpha e_1}{Nm + 1} + \frac{\alpha(\alpha - 1)}{2} \frac{e_1^2}{(Nm + 1)^2} + \frac{\alpha e_0 e_1}{Nm + 1} \right\} \right. \]

\[ \left. - w_2 \frac{\beta}{R} \left\{ e_1 + e_1 e_3 - e_1 e_2 + \frac{\alpha e_1^2}{Nm + 1} \right\} - 1 \right] \]

(2.3)

Taking expectation of both sides of (2.3), we get the bias of \( t \) to the first degree of approximation as:

\[ B(t) = \bar{Y} \left\{ (w_1 - 1) + w_1 \left\{ \frac{\alpha}{Nm + 1} f_1 C_{py} + \frac{\alpha(\alpha - 1)}{2(Nm + 1)^2} f_1 C_p^2 \right\} \right. \]

\[ \left. - w_2 \frac{\beta}{R} f_1 \left\{ C_p \frac{\lambda_{12}}{\rho_{pb}} - C_p \lambda_{03} + \frac{\alpha}{Nm + 1} C_p^2 \right\} \right\} \]

(2.4)

Squaring both sides of (2.3) and neglecting terms of \( e \)'s having power greater than two, we have
\[(t - Y)^2 = Y^2 \left[ 1 + 2e_0 + \frac{2\alpha e_1}{Nm + 1} + e_0^2 + \frac{\alpha(2\alpha - 1)e_1^2}{(Nm + 1)^2} + \frac{4\alpha e_0 e_1}{Nm + 1} \right]
+ \frac{w_2}{\beta} \left[ \frac{\beta^2}{R} e_1^2 + 1 - 2w_1w_2 \frac{\beta}{R} \left[ e_1 + e_1e_3 + e_0e_1 - e_1e_2 + \frac{2\alpha e_1^2}{Nm + 1} \right] \right]
- 2w_1 \left[ 1 + e_0 + \frac{\alpha e_1}{Nm + 1} + \frac{\alpha e_0 e_1}{Nm + 1} + \frac{\alpha(\alpha - 1)e_1^2}{2(Nm + 1)^2} \right]
+ 2w_2 \frac{\beta}{R} \left[ e_1 + e_1e_3 - e_1e_2 + \frac{\alpha e_1^2}{Nm + 1} \right]
\]

Taking expectation of both sides of (2.5), we get the MSE of \(t\) to the first degree of approximation as:

\[
\text{MSE}(t) = Y^2 \left[ 1 + w_1^2 A_{1(\alpha)}^{(m)} + w_2^2 A_2 - 2w_1w_2 A_{3(\alpha)}^{(m)} - 2w_1 A_{4(\alpha)}^{(m)} + 2w_2 A_{5(\alpha)}^{(m)} \right]
\] (2.6)

where,

\[
A_{1(\alpha)}^{(m)} = \left[ 1 + f_1 \left\{ c_y^2 + \frac{\alpha C_p^2}{Nm + 1} \left( \frac{2(\alpha - 1)}{Nm + 1} + 4k \right) \right\} \right]
\]

\[
A_2 = \left( \frac{\beta}{R} \right)^2 f_1 c_p^2
\]

\[
A_{3(\alpha)}^{(m)} = \frac{\beta}{R} f_1 \left[ \frac{C_p^2}{N + 1} \left\{ \frac{2\alpha}{Nm + 1} + k \right\} + \frac{C_p}{\rho_p} \lambda_{12} - C_p \lambda_{03} \right]
\]

\[
A_{4(\alpha)}^{(m)} = \left[ 1 + \frac{\alpha}{Nm + 1} f_1 \left\{ \frac{\alpha - 1}{2(Nm + 1)} + k \right\} \right]
\]

\[
A_{5(\alpha)}^{(m)} = \frac{\beta}{R} f_1 \left[ \frac{\alpha C_p^2}{Nm + 1} + \frac{C_p}{\rho_p} \lambda_{12} - C_p \lambda_{03} \right]
\]

where, \(k = \rho_p \frac{C_y}{C_p}\).

The MSE(t) is minimised for

\[
w_1 = \frac{\left( A_2 A_{4(\alpha)}^{(m)} - A_{3(\alpha)}^{(m)} A_{5(\alpha)}^{(m)} \right)}{\left( A_{1(\alpha)}^{(m)} A_2 - \left( A_{3(\alpha)}^{(m)} \right)^2 \right)} = w_{10}
\] (2.7)

\[
w_2 = \frac{\left( A_{3(\alpha)}^{(m)} A_{4(\alpha)}^{(m)} - A_{1(\alpha)}^{(m)} A_{5(\alpha)}^{(m)} \right)}{\left( A_{1(\alpha)}^{(m)} A_2 - \left( A_{3(\alpha)}^{(m)} \right)^2 \right)} = w_{20}
\] (2.8)
3. Members of the family of estimator of $t$ and their Biases and MSE

Table 3.1: Different members of the family of estimators of $t$

<table>
<thead>
<tr>
<th>Choice of scalars</th>
<th>$w_1$</th>
<th>$w_2$</th>
<th>$\alpha$</th>
<th>$m$</th>
<th>Estimator</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0$</td>
<td></td>
<td>$t_1 = \bar{y}$</td>
</tr>
<tr>
<td>$w_1$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0$</td>
<td></td>
<td>$t_2 = w_1 \bar{y}$</td>
</tr>
<tr>
<td>$w_1$</td>
<td>$0$</td>
<td>$\alpha$</td>
<td>$m$</td>
<td></td>
<td>$t_3 = w_1 \bar{y} \left( \frac{a}{Q} \right)^\alpha$</td>
</tr>
<tr>
<td>$w_1$</td>
<td>$0$</td>
<td>$\alpha$</td>
<td>$0$</td>
<td></td>
<td>$t_4 = w \bar{y} \left( \frac{p}{P} \right)^\alpha$</td>
</tr>
<tr>
<td>$1$</td>
<td>$0$</td>
<td>$-1$</td>
<td>$0$</td>
<td></td>
<td>$t_5 = \left( \frac{p}{P} \right)$,</td>
</tr>
<tr>
<td>$1$</td>
<td>$1$</td>
<td>$-1$</td>
<td>$0$</td>
<td></td>
<td>$t_6 = \left[ \bar{y} + b(P - p) \right] \frac{p}{P}$</td>
</tr>
<tr>
<td>$w_1$</td>
<td>$w_2$</td>
<td>$0$</td>
<td>$0$</td>
<td></td>
<td>$t_7 = \left[ w_1 \bar{y} + w_2 b(P - p) \right]$</td>
</tr>
<tr>
<td>$w_1$</td>
<td>$1$</td>
<td>$0$</td>
<td>$0$</td>
<td></td>
<td>$t_8 = \left[ w_1 \bar{y} + b(P - p) \right]$</td>
</tr>
<tr>
<td>$w$</td>
<td>$w$</td>
<td>$0$</td>
<td>$0$</td>
<td></td>
<td>$t_9 = \left[ \bar{y} + b(P - p) \right]$</td>
</tr>
<tr>
<td>$1$</td>
<td>$1$</td>
<td>$0$</td>
<td>$0$</td>
<td></td>
<td>$t_{10} = \bar{y} + b(P - p)$</td>
</tr>
</tbody>
</table>

The estimator $t_1 = \bar{y}$ is an unbiased estimator of the population mean $\bar{Y}$ and has the variance

$$\text{Var}(t_1) = f_1 S_{\bar{y}}^2$$

(3.1)
To, the first degree of approximation the biases and MSE’s of $t_i$’s, $i=1,2,\ldots,10$ are respectively given by

$$B(t_2) = \bar{Y}(w_1 - 1)$$

$$B(t_3) = \bar{Y}\left[(w_1 - 1) + f_1 w_1 \left\{ \frac{\alpha}{Nm+1} \rho_{pb} C_y C_p + \frac{\alpha(\alpha-1)}{2(Nm+1)} C_p^2 \right\} \right]$$

$$B(t_4) = \bar{Y}\left[(w_1 - 1) + w_1 f_1 \left\{ \alpha \rho_{pb} C_y C_p + \frac{\alpha(\alpha-1)}{2} C_p^2 \right\} \right]$$

$$B(t_5) = \bar{Y} f_1 \left[C_p^2 - \rho_{pb} C_y C_p \right]$$

$$B(t_6) = \bar{Y} f_1 \left[(C_p^2 - \rho_{pb} C_y C_p) - \frac{\beta}{R} \left(C_p \frac{\lambda_{12}}{\rho_{pb}} - C_p \lambda_{03} \right) \right]$$

$$B(t_7) = \bar{Y}\left[(w_1 - 1) - \frac{\lambda_{12} \beta}{R} f_1 \left\{ C_p \frac{\lambda_{12}}{\rho_{pb}} - C_p \lambda_{03} \right\} \right]$$

$$B(t_8) = \bar{Y}\left[(w_1 - 1) - \frac{\beta}{R} f_1 \left\{ C_p \frac{\lambda_{12}}{\rho_{pb}} - C_p \lambda_{03} \right\} \right]$$

$$B(t_9) = \bar{Y}\left[(w_1 - 1) - w f_1 \beta \frac{1}{R} \left\{ C_p \frac{\lambda_{12}}{\rho_{pb}} - C_p \lambda_{03} \right\} \right]$$

$$B(t_{10}) = -\bar{Y} \frac{\beta}{R} f_1 \left[C_p \frac{\lambda_{12}}{\rho_{pb}} - C_p \lambda_{03} \right]$$

The corresponding MSE’s will be

$$\text{MSE}(t_2) = \bar{Y}^2 \left[1 + w_1^2 A_{l(0)}^{(0)} - 2 w_1 A_{l(0)}^{(0)} \right]$$

$$\text{MSE}(t_3) = \bar{Y}^2 \left[1 + w_1^2 A_{l(0)}^{(m)} - 2 w_1 A_{l(0)}^{(m)} \right]$$

$$\text{MSE}(t_4) = \bar{Y}^2 \left[1 + w_1^2 A_{l(0)}^{(0)} - 2 w_1 A_{l(0)}^{(0)} \right]$$

$$\text{MSE}(t_5) = \bar{Y}^2 \left[1 + A_{l(-1)}^{(0)} - 2 A_{l(-1)}^{(0)} \right]$$
MSE(t₆) = \bar{Y}^{2} \left[ 1 + A^{(0)}_{1(-1)} + A_{2} - 2A^{(0)}_{3(-1)} - 2A^{(0)}_{4(-1)} + 2A^{(0)}_{5(-1)} \right] (3.15)

MSE(t₇) = \bar{Y}^{2} \left[ 1 + w_{1}^{2}A^{(0)}_{1(0)} + w_{2}^{2}A_{2} - 2w_{1}w_{2}A^{(0)}_{3(0)} - 2w_{1}A^{(0)}_{4(0)} + 2w_{2}A^{(0)}_{5(0)} \right] (3.16)

MSE(t₈) = \bar{Y}^{2} \left[ 1 + w_{1}^{2}A^{(0)}_{1(0)} + A_{2} - 2w_{1}\left(A^{(0)}_{3(0)} + A^{(0)}_{4(0)}\right) + 2A^{(0)}_{5(0)} \right] (3.17)

MSE(t₉) = \bar{Y}^{2} \left[ 1 + w_{1}^{2}\left(A^{(0)}_{1(0)} + A_{2} - 2A^{(0)}_{4(0)}\right) - 2w\left(A^{(0)}_{4(0)} - A^{(0)}_{5(0)}\right) \right] (3.18)

MSE(t₁₀) = \bar{Y}^{2} \left[ 1 + A^{(0)}_{1(0)} + A_{2} - 2\left(A^{(0)}_{3(0)} + A^{(0)}_{4(0)} - A^{(0)}_{5(0)}\right) \right] (3.19)

The MSE’s of the estimators of \( t_i \), \( i = 2, 3, 4, 7, 8, 9 \) will be minimised respectively, for

\[ w_{1} = \frac{A^{(0)}_{4(0)}}{A^{(0)}_{1(0)}} \] (3.20)

\[ w_{1} = \frac{A^{(m)}_{4(\alpha)}}{A^{(m)}_{1(\alpha)}} \] (3.21)

\[ w_{1} = \frac{A^{(0)}_{4(\alpha)}}{A^{(0)}_{1(\alpha)}} \] (3.22)

\[ w_{1} = \left( A_{2}A^{(0)}_{4(0)} - A^{(0)}_{3(0)}A^{(0)}_{5(0)} \right) \left( A_{2}A^{(0)}_{3(0)} - A^{(0)}_{3(0)} \right)^{2} \] (3.23)

\[ w_{2} = \frac{\left( A^{(0)}_{3(0)}A^{(0)}_{4(0)} - A^{(0)}_{1(0)}A^{(0)}_{5(0)} \right)}{A_{2}A^{(0)}_{1(0)} - \left( A^{(0)}_{3(0)} \right)^{2}} \]

\[ w_{1} = \frac{\left( A^{(0)}_{3(0)} + A^{(0)}_{4(0)} \right)}{A^{(0)}_{1(0)}} \] (3.24)

\[ w = \frac{A^{(0)}_{4(0)} - A^{(0)}_{3(0)}}{A^{(0)}_{1(0)} + A_{2} - 2A^{(0)}_{3(0)}} \] (3.25)

Thus the resulting minimum MSE of \( t_i \), \( i = 2, 3, 4, 7, 8, 9 \) are, respectively given by
min. $\text{MSE}(t_2) = \overline{Y}^2 \left[ 1 - \frac{(A_{4(0)})^2}{A_{1(0)}} \right]$ \hfill (3.26)

min. $\text{MSE}(t_3) = \overline{Y}^2 \left[ 1 - \frac{(A_{m})^{2}}{A_{l}(\alpha)} \right]$ \hfill (3.27)

min. $\text{MSE}(t_4) = \overline{Y}^2 \left[ 1 - \frac{(A_{4(\alpha)})^2}{A_{1(\alpha)}} \right]$ \hfill (3.28)

min. $\text{MSE}(t_7) = \overline{Y}^2 \left[ 1 - \frac{A_2 \left( A_{4(0)} \right)^2 - 2A_{3(0)}A_{5(0)} + A_{1(0)} \left( A_{5(0)} \right)^2}{A_2 A_{1(0)}^2 - A_{3(0)}^2} \right]$ \hfill (3.29)

min. $\text{MSE}(t_8) = \overline{Y}^2 \left[ 1 + A_2 + 2A_{5(0)} - \frac{\left( A_{3(0)} - A_{4(0)} \right)^2}{A_{1(0)}} \right]$ \hfill (3.30)

min. $\text{MSE}(t_9) = \overline{Y}^2 \left[ 1 - \frac{(A_{4(0)} - A_{5(0)})^2}{A_{1(0)} + A_2 - 2A_{3(0)}} \right]$ \hfill (3.31)

4. **Empirical study**

The data for the empirical study is taken from natural population data set considered by Sukhatme and Sukhatme (1970):

\[
y = \text{Number of villages in the circles and } \\
\phi = \text{A circle consisting more than five villages}
\]

\[
N = 89, \overline{Y} = 3.36, P = 0.1236, \rho_{pb} = 0.766, C_y = 0.6040, C_p = 2.190
\]

\[
\lambda_{04} = 6.1619, \lambda_{40} = 3.810, \lambda_{12} = 146.475, \lambda_{03} = 2.2744
\]

In the Table 4.1 percent relative efficiencies (PRE’s) of various estimators are computed with respect to $\overline{Y}$. 

---

Rajesh Singh ■ Florentin Smarandache (editors)
Table 4.1: PRE of different estimators of $\bar{Y}$ with respect to $\bar{y}$.

<table>
<thead>
<tr>
<th>Estimator</th>
<th>PRE’s</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_1 = \bar{y}$</td>
<td>100.00</td>
</tr>
<tr>
<td>$t_2$</td>
<td>101.41</td>
</tr>
<tr>
<td>$t_3$</td>
<td>90.35</td>
</tr>
<tr>
<td>$t_4$</td>
<td>6.92</td>
</tr>
<tr>
<td>$t_5$</td>
<td>11.64</td>
</tr>
<tr>
<td>$t_6$</td>
<td>7.38</td>
</tr>
<tr>
<td>$t_7$</td>
<td>100.44</td>
</tr>
<tr>
<td>$t_8$</td>
<td>243.39</td>
</tr>
<tr>
<td>$t_9$</td>
<td>243.42</td>
</tr>
<tr>
<td>$t_{10}$</td>
<td>241.98</td>
</tr>
</tbody>
</table>

**Conclusion**

The MSE values of the members of the family of the estimator $t$ have been obtained using (2.6). These values are given in Table 4.1. When we examine Table 4.1, we observe the superiority of the proposed estimators $t_2$, $t_7$, $t_8$, $t_9$ and $t_{10}$ over usual unbiased estimator $t_1$, $t_3$, $t_4$, Naik and Gupta (1996) estimator $t_5$ and Singh et. al. (2008) estimator $t_6$. From this result we can infer that the proposed estimators $t_8$ and $t_9$ are more efficient than the rest of the estimators considered in this paper for this data set.

We would also like to remark that the value of the min. MSE($t_{10}$), which is equal to the value of the MSE of the regression estimator is 241.98. From Table 4.1 we notice that the value of MSE of the estimators $t_8$ and $t_9$ are less than this value, as shown in Table 4.1. Finally, we can say that the proposed estimators $t_8$ and $t_9$ are more efficient than the regression estimator for this data set.

**References**


Estimation of Ratio and Product of Two Population Means Using Auxiliary Characters in the Presence of Non Response

B. B. Khare
Department of Statistics, Banaras Hindu University, Varanasi (U.P), India
Email: bbkhare56@yahoo.com

Abstract
The auxiliary information is used in increasing the efficiency of the estimators for the parameters of the populations such as mean, ratio, and product of two population means. In this context, the estimation procedure for the ratio and product of two population means using auxiliary characters in special reference to the non response problem has been discussed.

Keywords Auxiliary variable, MSE, non response, SRS, efficiency.

Introduction
The use of auxiliary information in sample surveys in the estimation of population mean, ratio, and product of two population means has been studied by different authors by using different estimation procedures. The review work in this topic has been given by Tripathi et al. (1994) and Khare (2003). In the present context the problems of estimation of ratio and product of two population means have been considered in different situations especially in the presence of non response.

Estimation of Ratio and product of two population means

Case 1. The Case of Complete Response:
Singh (1965,69), Rao and Pareira (1968), Shahoo and Shahoo (1978), Tripathi (1980), Ray and Singh (1985) and Khare (1987) have proposed estimators of ratio and product of two population means using auxiliary characters with known mean. Singh (1982) has proposed the case of double sampling for the estimation of ratio and product of two population mean. Khare (1991(a)) has proposed a class of estimators for R and P using double sampling scheme, which are given as follows:

\[ R^* = f(v,u) \quad \text{and} \quad P^* = g(w,u) \]
such that \( f(R,1) = R, \ g(P,1) = P, \ f_1(R,1) = 1 \) and \( g_1(P,1) = 1 \), where \( v = \left( \frac{\bar{y}_1}{\bar{y}_2} \right), \ w = \bar{y}_1 \bar{y}_2 \) and \( u = \left( \frac{x_1'}{x_1} \right) \). Here \( \bar{y}_1, \bar{y}_2 \) and \( \bar{x}_1 \) denote the sample mean of study characters \( y_1, y_2 \) and auxiliary character \( x_1 \) based on a sub sample of size \( n ( < n') \) and \( \bar{x}_1' \) is sample mean of \( x_1 \) based on a larger sample of size \( n' \) drawn by using SRSWOR method of sampling from the population of size \( N \). The first partial derivatives of \( f(v,u) \) and \( g(w,u) \) with respect to \( v \) and \( w \) are denoted by \( f_1(v,u) \) and \( g_1(w,u) \) respectively. The function \( f(v,u) \) and \( g(w,u) \) also satisfied some regularity conditions for continuity and existence of the functions. The sample size for first phase and second phase sample which may be from the first phase sample or independent of first phase sample drawn from the remaining part of the population \( (N - n') \).

Singh et al. (1994) have extended the class of estimators proposed by Khare (1991(a)) and proposed a new class of estimator for \( R \), which is given as follows:

\[
R_g = \hat{R} h(u', v')
\]

(2)

where \( \hat{R} = \frac{\bar{y}_1}{\bar{y}_2}, \ u' = \frac{\bar{x}}{\bar{x}'}, \) and \( v' = \frac{s_x^2}{s_x'^2} \), where \( (\bar{x}, s_x^2) \) and \( (\bar{x}', s_x'^2) \) are sample mean and sample mean square of auxiliary character based on \( n \) and \( n' (> n) \) units respectively.

Srivastava et al. (1988,89) have suggested chain ratio estimators for \( R \) and \( P \). Which are given as follows:

\[
R'_1 = \hat{R} \left( \frac{\bar{y}_3}{\bar{y}_3}\right)^{\alpha_1} \left( \frac{\bar{y}_4}{\bar{y}_4}\right)^{\alpha_2} \quad \text{and} \quad R'_2 = \hat{R} \left( \frac{\bar{y}_3}{\bar{y}_3}\right)^{\beta_1} \left( \frac{\bar{y}_4}{\bar{y}_4}\right)^{\beta_2}
\]

(3)

\[
P'_1 = \hat{P} \left( \frac{\bar{y}_3}{\bar{y}_3}\right)^{\alpha_1} \left( \frac{\bar{y}_4}{\bar{y}_4}\right)^{\alpha_2} \quad \text{and} \quad P'_2 = \hat{P} \left( \frac{\bar{y}_3}{\bar{y}_3}\right)^{\beta_1} \left( \frac{\bar{y}_4}{\bar{y}_4}\right)^{\beta_2}
\]

(4)

Further Singh et al. (1994) have given a general class of estimators

\[
\hat{R}_h = h(\hat{R}, u, v) \quad \text{and} \quad \hat{P}_h = h(\hat{P}, u, v),
\]

(5)

such that \( h(R,1,1) = R \) and \( h(P,1,1) = P \), where \( u = \left( \frac{\bar{y}_3}{\bar{y}_3} \right) \) and \( v = \left( \frac{\bar{y}_4}{\bar{y}_4} \right) \). The functions \( h(\hat{R}, u, v) \) and \( h(\hat{P}, u, v) \) satisfy the regularity conditions.

Khare (1991(b)) have proposed the class of estimators for using multi-auxiliary characters with known means. which are given as follows:

\[
R' = \hat{R} h(u_1, u_2, .., u_p) = \hat{R} h(u) \quad \text{and} \quad R'' = g(\hat{R}, u),
\]

(6)
such that \( h(e) = 1 \) and \( g(\hat{R}, e) = R \), where \( h(u) \) and \( g(\hat{R}, u) \) satisfying some responding conditions.

Further, Khare (1993(a)) has proposed a class of estimators for \( R \) using multi-auxiliary characters with unknown means, the class of estimators is given as follows:

\[
R'_m = g(\hat{R}, u'),
\]

such that \( g(\hat{R}, e) = 1 \), where \( u'_i = \frac{x'_i}{\bar{x}_j} \), \( u' = (u_1, u_2, \ldots, u_p) \), \( \bar{x}_j \) and \( x'_i \) are sample mean based on \( n \) and \( n' (> n) \) units for auxiliary characters \( x_i, i = 1, 2, \ldots, p \).

Similarly, Khare (1992) have proposed class of estimators for \( P \) using \( p \) auxiliary characters with known and unknown population mean and studied their properties.

Further, Khare (1990) has proposed a generalized class of estimator for a combination of product and ratio of some population means using multi-auxiliary characters. The parametric combination is given by:

\[
\theta = \frac{\bar{Y}_1, \bar{Y}_2, \bar{Y}_3, \ldots, \bar{Y}_m}{\bar{Y}_{m+1}, \bar{Y}_{m+2}, \bar{Y}_{m+3}, \ldots, \bar{Y}_k},
\]

which is the product of first \( m \) population means \( \bar{Y}_1, \bar{Y}_2, \bar{Y}_3, \ldots, \bar{Y}_m \) divided by product of \( k - m \) population means \( \bar{Y}_{m+1}, \bar{Y}_{m+2}, \bar{Y}_{m+3}, \ldots, \bar{Y}_k \) respectively. The conventional estimator for \( \theta \) is given by

\[
\hat{\theta} = \frac{\bar{Y}_1, \bar{Y}_2, \bar{Y}_3, \ldots, \bar{Y}_m}{\bar{Y}_{m+1}, \bar{Y}_{m+2}, \bar{Y}_{m+3}, \ldots, \bar{Y}_k},
\]

It is important to note that for \( m = 1, k = 2; \quad \theta = R \)

\( m = 2, k = 2; \quad \theta = P \)

\( m = 1, k = 1; \quad \theta = \bar{Y}_1 \)

\( m = k = 1; \quad \theta = \bar{Y}_1 \)

\( m = k = 3; \quad \theta = \bar{Y}_1 \)

\( m = 2, k = 4; \quad \theta = \bar{Y}_1 \)

Using \( p \) auxiliary characters \( x_1, x_2, \ldots, x_p \) with known population means \( \bar{X}_1, \bar{X}_2, \ldots, \bar{X}_p \) the class of estimators \( \theta^* \) is given by:

\[
\theta^* = \hat{h}(u),
\]
such that \( h(\varepsilon) = 1 \), where \( u' = (u_1, u_2, \ldots, u_p) \) and \( u_i = \frac{x_i}{\bar{x}_i}, i = 1, 2, \ldots, p \).

The function \( h(u_1, u_2, \ldots, u_p) = h(\mathbf{u}) \) satisfied the following regularity conditions:

a) Whatever be the sample chosen \( \{\mathbf{u}\} \), assume values in a bounded closed convex subset \( G \) of \( p \)-dimensional real space containing the point \( \mathbf{u} = \varepsilon \).

b) In \( G \), the function \( h(\mathbf{u}) \) is continuous and bounded.

c) The first and second partial derivatives of \( h(\mathbf{u}) \) exists and are continuous and bounded in \( G \).

For two auxiliary variables it is found that the lower bond of the variance of the class of estimators \( \theta^* \) is same as given by the estimators proposed by Singh (1969) and Shah and Shah (1978). Hence it is remarked that the class of estimators \( \theta^* \) will attain lower bound for mean square error if the specified and regularity conditions are satisfied.

Further, Khare (1993b) have proposed the class of two phase sampling estimators for the combination of product and ratio of some population means using multi-auxiliary characters with unknown population means, which is given as follows:

\[ \theta^* = \hat{\theta}(\mathbf{y}), \]

where \( \mathbf{y} = (y_1, y_2, \ldots, y_p)' \), \( y_i = \frac{\bar{y}_i}{\bar{x}_i}, i = 1, 2, \ldots, p \).

Such that \( h(\varepsilon) = 1 \) and \( h(\mathbf{u}) \) satisfies some regularly conditions.

**Case 2. Incomplete Response in the Sample due to Non-response:**

In case of non-response on some units selected in the sample, Hansen and Hurwitz (1946) have suggested the method of sub sampling from non-respondents and proposed the estimator for population mean. Further, Khare et al. (2014) have proposed some new estimators in this situation of sub sampling from non-respondents.

Khare & Pandey (2000) and Khare & Sinha (2010) have proposed the class of estimators for ratio and product of two population means using auxiliary character with known population mean in the presence of non-response on the study characters, which is given as follows:

\[ R_i^* = R^* h(u_i) \quad \text{and} \quad P_i^* = P^* h(u_i), i = 1, 2, \]

such that \( h(1) = 1 \), where \( R^* = \frac{\bar{y}_1}{\bar{y}_2}, \quad P^* = \frac{\bar{y}_1\bar{y}_2}{\bar{x}}, \quad u_1 = \frac{\bar{x}}{\bar{x}}, u_2 = \frac{\bar{x}}{\bar{x}} \) and \( \bar{y}_1, \bar{y}_2 \) and \( \bar{x} \) are sample means for \( y_1, y_2 \) and \( x \) characters proposed by Hansen and Hurwitz (1946) based on \( n_1 + r \) units and \( \bar{x} \) is the sample mean based on \( n \) units. Khare & Sinha (2012) have proposed a combined class of estimators for ratio and product of two population mean in the presence of non-response with known population mean \( \bar{x} \). This is a more general class of estimators for \( R \) and \( P \) under some specified and regularity conditions. Khare et al. (2013 (a)) have proposed an improved class of estimators for \( R \). In this case, the improved class of estimators for \( R \) using auxiliary character with known population mean \( \bar{x} \) in the presence of non response is given as follows:
such that \( g(R, 1) = R \), \( g_1(R, 1) \) and \( g_{12}(R, 1) = R^{-1} g_2(R, 1) \), where \( v = \frac{\hat{y}_1}{\hat{y}_2} \), \( u_1 = \hat{x} \) and \( u_2 = \hat{y} \). The function \( g(v, u_i) \) is assumed to be continuous and bounded in a real line and its first and second order partial derivatives exists. The first partial derivative of \( g(v, u_i) \) at the point \((R, 1)\) with respect to \( v \) and \( u_i \) is denoted by \( g_1(R, 1) \) and \( g_2(R, 1) \). The second order partial derivative of \( g(v, u_i) \) with respect to \( v, v \), and \( u_i \) at the point \((R, 1)\) is denoted by \( g_{11}(R, 1) \), \( g_{12}(R, 1) \) and \( g_{22}(R, 1) \) respectively. Some members of the class of estimators \( R_i \) are given as follows:

\[
C_1 = w_0 v u_i^{\alpha_i}, \quad C_2 = w_1 v + w_2 u_i, \quad C_3 = w_1' v + w_2' u_i^{\beta_i}, \quad i = 1, 2,
\]

where \( w_0, w_1, w_2, w_1', w_2', \alpha_i \) and \( \beta_i \) \( i = 1, 2 \) are constants. Further the class of estimator proposed by Khare and Sinha (2013) is more efficient than the estimator proposed by Khare and Pandey (2000).

Further, Khare and Sinha (2002(a, b)) have proposed two phase sampling estimators for ratio and product of two population means in the presence of non-response. Khare and Sinha (2004(a,b)) have proposed a more general class of two phase sampling estimators for \( R \) and \( P \). which are given as follows:

\[
T_i = g(v, u_i), \quad i = 1, 2,
\]

such that \( g(R, 1) = R \) and \( g_1(R, 1) = 1 \), where \( v = \frac{\hat{y}_1}{\hat{y}_2} \), \( u_1 = \frac{\hat{x}}{\hat{y}} \), \( u_2 = \frac{\hat{x}}{\hat{y}} \) and \( \hat{y} \) is sample mean based on \( n' (> n) \) units. The function \( g(v, u_i) \) satisfy some regularly conditions.

\[
T_i^* = g(w, u_i), \quad i = 1, 2,
\]

such that \( g(P, 1) = 1 \) and \( g_1(P, 1) = 1 \), where \( w = \frac{\hat{y}_1}{\hat{y}} \), \( u_1 = \frac{\hat{x}^*}{\hat{y}^*} \), \( u_2 = \frac{\hat{x}^*}{\hat{y}^*} \) and \( g(w, u_i) \) satisfy some regularly conditions.

Khare et al. (2012) have proposed two generalized chain type estimators \( T_{g1} \) and \( T_{g2} \) for \( R \) using auxiliary characters in the presence of non-response, which are given as follows:

\[
T_{g1} = \hat{R} \left( \frac{\hat{x}^*}{\hat{y}} \right)^{\alpha_1} \left( \frac{\hat{y}^*}{\hat{y}} \right)^{\alpha_2} \quad \text{and} \quad T_{g2} = \hat{R} \left( \frac{\hat{x}}{\hat{y}} \right)^{\beta_1} \left( \frac{\hat{y}}{\hat{y}} \right)^{\beta_2}
\]

where \( \hat{R} = \frac{\hat{y}_1}{\hat{y}_2} \) and \( (\alpha_1, \alpha_2) \) and \( (\beta_1, \beta_2) \) are suitable constants. It has been observed that due to use of additional auxiliary character with known population mean along with the main auxiliary character, the proposed class of estimators \( T_{g1} \) and \( T_{g2} \) are more efficient than the
corresponding generalized estimators for $R$ using the main auxiliary character only in the case of two phase sampling in the presence of non response for fixed sample sizes ($n’, n$) and also for fixed cost ($C \leq C_0$). It is also seen that less cost is incurred for $T_{g1}$ and $T_{g2}$ than the cost incurred in the generalized estimator for $R$ in the case of two phase sampling in the presence of non response for specified precision ($V = V_0$).

Further, generalized chain estimators for ratio and product of two population means have been improved by putting $R = k_i \hat{R}$ and $P = k_i \hat{P}$ in place of $\hat{R}$ and $\hat{P}$ in the proposed estimators of $R$ and $P$. Further, Khare et al. (2013 (b)) have proposed the improved class of chain type estimators for ratio of two population means using two auxiliary characters in the presence of non-response. The class of estimators is given as follows:

$$R_{\alpha} = f(\hat{R}, u_i, v), \quad i = 1, 2, \quad (18)$$

such that $f(R, 1, 1) = 1$ and $f_1(R, 1, 1) = 1$, where $\hat{R} = \frac{x_i}{y_2}, \quad u_1 = \frac{x_i}{x'}, \quad u_2 = \frac{x}{x'}$ and $v = \frac{x'}{Z}$. The function $f(\hat{R}, u_i, v), \quad i = 1, 2$ satisfies some regularity conditions.

Khare and Sinha (2007) have proposed estimator for $R$ using multi-auxiliary characters with known population mean in the presence of non-response. The class of estimators $t_i$ is given as follows:

$$t_i = \hat{R} g_i(u_i’), \quad i = 1, 2, \quad (19)$$

such that $g_i(e_i’) = 1$, where $u_i$ and $e_i$ denote the column vectors $(u_{i1}, u_{i2}, ..., u_{ip})’$ and $(1, 1, ..., 1)’, \quad u_{i1} = \frac{x_i}{x_j} \quad \text{and} \quad u_{i2} = \frac{x_j}{x_i} \quad \text{for} \quad j = 1, 2, ..., p.

An improved under class of estimators for $R$ using multi-auxiliary variables using double sampling scheme in the presence of non-response has been proposed by Khare and Sinha (2012) and studies their properties.

Khare and Sinha (2014) have extended the class of estimator proposed by Khare and Sinha (2012) and proposed a wider class of two phase sampling estimators for $R$ using multi-auxiliary characters in the presence of non-response.

References


On The Use of Coefficient of Variation and $\beta_1, \beta_2$ in Estimating Mean of a Finite Population

$^1$B. B. Khare, $^1$P. S. Jha and $^2$U. Srivastava

$^1$Department of Statistics, B.H.U, Varanasi-221005
$^2$Statistics Section, MMV, B.H.U, Varanasi-221005
bbkhare56@yahoo.com

Abstract
In this paper the use of coefficient of variation and shape parameters in each stratum, the problem of estimation of population of mean has been considered. The expression of mean squared error of the proposed estimator is derived and its properties are discussed.

Keywords Auxiliary information, MSE, coefficient of variation, stratum, shape parameter.

Introduction
The use of prior information about the population parameters such as coefficient of variation, mean and skewness and kurtosis are very useful in the estimation of the population parameter of the study character. In agricultural and biological studies information about the coefficient of variation and the shape parameters are often available. If these parameters remain essentially unchanged over the time than the knowledge about them in such case it may profitably be used to produce optimum estimates of the parameters (Sen and Gerig (1975)). Searls (1964, 67) and Hirano (1972) have proposed the use of coefficient of variation in the estimation the population mean. Searl and Intarapanich (1990) have suggested the use of kurtosis in the estimation of variance. Sen (1978) has proposed the estimator for population mean using the known value of coefficient of variation.

In Stratified random sampling, the theory has been developed to provide the optimum estimator $T_1$ of the population mean based on sample mean from each stratum. We extend it by constructing an estimator $T_2$ using the coefficient of variation $C_i$ and shape parameter $\beta_{1i}, \beta_{2i} (i = 1, 2, \ldots, K)$ from each stratum and discuss its usefulness. We also define estimators $T_3$ and $T_4$ when the coefficients of variation are unknown but shape parameters are known and when neither the coefficients of variation are known nor the shape parameters are known.

Estimators and their Mean Square Error
Let $N_i$ denotes the size of the $i^{th}$ stratum and $n_i$ denotes the size of the sample to be selected from the $i^{th}$ stratum and $h$ be the number of strata with
\[ \sum_{i=1}^{h} N_i = N \quad \text{and} \quad \sum_{i=1}^{h} n_i = n, \quad (1) \]

where \( N \) and \( n \) denote the number of units in the population and sample respectively.

Let \( y_{ij} \) be the \( j^{th} \) unit of the \( i^{th} \) stratum. Then the population mean \( \overline{Y}_N \) can be expressed as

\[ \overline{Y}_N = \sum_{i=1}^{h} p_i \overline{Y}_i, \quad (2) \]

where \( p_i = \frac{N_i}{N} \) and \( \overline{Y}_i \) is the population mean for the \( i^{th} \) stratum.

Let \( n_i \) units be selected from the \( i^{th} \) stratum and the corresponding sampling mean and sample variance be denoted by \( \overline{y}_i \) and \( s_i^2 \) respectively. Then the estimate of \( \overline{Y}_N \) is given by

\[ T_1 = \sum_{i=1}^{h} p_i \overline{y}_i \quad (3) \]

and the

\[ V(T_1) = \sum_{i=1}^{h} \frac{p_i^2 \sigma_i^2}{n_i} \quad \text{(if f.p.c is ignored),} \quad (4) \]

where \( \sigma_i^2 \) is the population variance of \( y \) in the \( i^{th} \) stratum.

**Case 1: Coefficient of variation and the shape parameters are known.**

We defined

\[ T_2 = \sum_{i=1}^{h} p_i \{ \alpha_i \overline{y}_i + (1 - \alpha_i)C_i^{-1} \sqrt{s_i^2} \} \quad (5) \]

and expectation of \( T_2 \) is given by

\[ E(T_2) = \sum_{i=1}^{h} p_i \{ \alpha_i \overline{y}_i + (1 - \alpha_i)(1 - \frac{1}{8} \frac{V(s_i^2)}{\sigma_i^4}) \} \]

\[ = \sum_{i=1}^{h} p_i \{ \overline{y}_i - (1 - \alpha_i)\frac{1}{8} \frac{V(s_i^2)}{\sigma_i^4}) \} \]

\[ = \sum_{i=1}^{h} p_i \{ \overline{y}_i - \frac{(1 - \alpha_i)}{8} \left( \frac{\beta_{2i} - 1}{n_i} + \frac{2}{n(n-1)} \right) \} \]

\[ = \overline{Y}_N + O \left( \frac{1}{n_i} \right) \]
where $\beta_{2i}$ is the measure of kurtosis in the $i^{th}$ stratum.

The bias in $T_2$ is of order $\frac{1}{n_i}$ and will be negligible for large $n_i$'s.

The mean square error of the estimator is

$$
\text{MSE}(T_2 / \alpha_i) = \frac{k}{n_i} \sum p_i^2 \sigma_i^2 \left[ \alpha_i^2 + \alpha_i(1 - \alpha_i)C_i^{-4} \sqrt{\beta_{li}} + \frac{(1 - \alpha_i)^2}{4} C_i^{-2}(\beta_{2i} - 1) \right] + O\left(\frac{1}{n_i^{3/2}}\right). \tag{7}
$$

Minimising (7) with respect to $\alpha_i$, we get the optimum value of $\alpha_i$ is given by

$$
\alpha_{i opt} = \frac{\beta_{2i} - 2C_i \sqrt{\beta_{li}} - 1}{4C_i^2 + 4C_i \sqrt{\beta_{li}} + \beta_{2i} - 1}, \tag{8}
$$

where $\beta_{li}$ is the measure of kurtosis in the $i^{th}$ stratum.

On putting the optimum value of $\alpha_{i opt}$ from (8) in (7) and on simplification we get

$$
\text{MSE}(T_2)_{\text{min}} = \frac{k}{n_i} \sum p_i^2 \sigma_i^2 \left[ \frac{\beta_{2i} - \beta_{li} - 1}{(\beta_{2i} - \beta_{li} - 1) + (\sqrt{\beta_{li}} - 2C_i)^2} \right] + O\left(\frac{1}{n_i^{3/2}}\right). \tag{9}
$$

The value of $\alpha_{i opt}$ will be less than one for $\sqrt{\beta_{li}} < 2C_i$, which implies that the distribution is near normal, poison, negative binomial and Neyman type I. The value of $\alpha_{i opt}$ will be equal to one for $\sqrt{\beta_{li}} = 2C_i$, which is true for gamma and exponential distribution. The value of $\alpha_{i opt}$ will be greater than one for $\sqrt{\beta_{li}} > 2C_i$, which is likely to the distribution of lognormal or inverse Gaussian. It is easy to see that $T_2$ will always be more efficient than $T_1$ if $\sqrt{\beta_{li}} < 2C_i$ or $\sqrt{\beta_{li}} > 2C_i$, justifying the use of $T_2$ in the case of near normal, poison, negative binomial, Neyman type I and lognormal or inverse Gaussian distribution. $T_2$ is equally efficient $T_1$, if $\sqrt{\beta_{li}} = 2C_i$ and so for in gamma or exponential distribution one may use $T_1$ or $T_2$. This shows that proposed estimator $T_2$ is uniformly superior to the estimator $T_1$, though a comparatively high efficiency may be seen in near normal, poison, negative binomial than lognormal or inverse Gaussian distribution.
Case 2: $C_i$’s are unknown, $\beta_{i1}$’s and $\beta_{2i}$’s are known.

When $C_i$’s are unknown, we use their estimates $c_i$ based on a larger sample of size $n_i'$ from a previous occasion. Now we define an estimator $T_3$ for $\bar{Y}_N$ given by

$$T_3 = \frac{1}{h} \sum_{i=1}^{h} p_i \{\alpha_i' \bar{y}_i + (1 - \alpha_i')c_i^{-1} \sqrt{s_i^2}\}$$  \hspace{1cm} (10)

The mean square error of the estimator $T_3$ as given by

$$\text{MSE}(T_3 / \alpha_i') = \frac{1}{h} \sum_{i=1}^{h} \frac{p_i^2 \sigma_i^2}{n_i'} \left[ \alpha_i'^2 + \alpha_i'(1 - \alpha_i')C_i^{-1} \sqrt{\beta_{ii} + (1 - \alpha_i')^2 C_i^{-2}} \{\beta_{2i} - 1\} + 4n_iC_i^{-2}V(c_i)\right].$$  \hspace{1cm} (11)

where $V(c_i) = \frac{C_i^2}{n_i'} \{\beta_{2i} - 1\} - \frac{\mu_3}{\mu_2^2 \mu_1'} + \frac{\mu_2}{\mu_1'^2} \leq \frac{C_i^2}{n_i'} (\beta_{2i} - 1) .

The optimum value of $\alpha_i'$ is given by

$$\alpha_{i\text{opt}} = \frac{(\beta_{2i} - \beta_{ii} - 1) + 4n_iC_i^{-2}V(c_i)}{(\beta_{2i} - \beta_{ii} - 1 + 4n_iC_i^{-2}V(c_i)) + (\sqrt{\beta_{ii} - 2C_i})^2} .$$  \hspace{1cm} (12)

It is easy to see that

$$\text{MSE}(T_3 / \alpha_{i\text{opt}})_{\text{min}} = \frac{1}{h} \sum_{i=1}^{h} \frac{p_i^2 \sigma_i^2}{n_i'} \left[ \frac{1}{(\beta_{2i} - \beta_{ii} - 1 + 4n_iC_i^{-2}V(c_i)) + (\sqrt{\beta_{ii} - 2C_i})^2} \right].$$  \hspace{1cm} (13)

It may be remarked that (13) differs from (9) by a single term $4n_iC_i^{-2}V(c_i)$ both in numerator and denominator. The nature of the estimator $T_3$ is similar to $T_2$ and its MSE will converge to $\text{MSE}(T_2)$ for $\frac{V(c_i)}{C_i^2} \rightarrow 0$.

Case 3: $C_i$’s, $\beta_{ii}$’s and $\beta_{2i}$’s are unknown:

When $C_i$’s, $\beta_{ii}$’s and $\beta_{2i}$’s are not known then they can be estimated on the basis of a larger sample of size $n_i'$ $\gg ...n_i$ from the past data and we may have the estimator for the population mean $\bar{Y}_N$ given by

$$T_4 = \frac{1}{h} \sum_{i=1}^{h} p_i \{\hat{\alpha}_i \bar{y}_i + (1 - \hat{\alpha}_i)c_i^{-1} \sqrt{s_i^2}\} ,$$  \hspace{1cm} (14)
where $\hat{\alpha}_{i, opt} = \frac{\hat{\beta}_{2i} - 2\hat{C}_i \sqrt{\hat{\beta}_{ll} - 1}}{4\hat{C}_i^2 + 4\hat{C}_i \sqrt{\hat{\beta}_{ll} + \hat{\beta}_{2i} - 1}}$.

It is easy to see that the $\text{MSE}(T_4)$ will be same as $\text{MSE}(T_3)$ because after estimating the unknown parameters in the constant $\alpha_{opt}$, the MSE will remains unchanged up to the terms of $O(n^{-1})$ (Srivastava and Jhajj (1983)).

References
A Study of Improved Chain Ratio-cum-Regression type Estimator for Population Mean in the Presence of Non-Response for Fixed Cost and Specified Precision

B. B. Khare, Habib Ur Rehman and U. Srivastava

Department Of Statistics, B.H.U, Varanasi-221005

Statistics Section, MMV, B.H.U, Varanasi-221005

Abstract

In this paper, a study of improved chain ratio-cum regression type estimator for population mean in the presence of non-response for fixed cost and specified precision has been made. Theoretical results are supported by carrying out one numerical illustration.

Keywords Simple random sampling, non response, fixed cost, precision.

Introduction

In the field of socio, economics, researches and agricultures the problem arises due to non-response which friendly occur due to not at home, lack of interest, call back etc. In this expression a procedure of sub sampling from non respondents was suggested by Hansen and Hurwitz (1946). The use of auxiliary information in the estimators of the population parameters have helped in increased the efficiency of the proposed estimator. Using auxiliary character with known population mean of the estimators have been proposed by Rao (1986,90) and Khare and Srivastava (1996,1997). Further, Khare and Srivastava (1993,1995),Khare et al. (2008), Singh and Kumar (2010), Khare and Kumar (2009) and Khare and Srivastava(2010) have proposed different types of estimators for the estimation of population mean in the presence of non-response in case of unknown population mean of the auxiliary character.

In the present paper, we have studied an improved chain ratio-cum-regression type estimator for population mean in the presence of non-response have proposed by Khare and Rehman (2014) in the case of fixed cost and specified precision. In the present study we have obtained the optimum size of first phase sample \( n' \) and second phase sample \( n \) is drawn from the population of size \( N \) by using SRSWOR method of sampling in case of fixed cost and also in case of specified precision \( V = V_0 \). The expression for the minimum MSE of the estimator has been obtained for the optimum values of \( n' \) and \( n \) in case of fixed cost \( C \leq C_0 \). The expression for minimum cost for the estimator has also been obtained in
case of specified precision \( V = V_0 \). An empirical study has been considered to observe the properties of the estimator in case of fixed cost and also in case of specified precision.

**The Estimators**

Let \( \bar{Y}, \bar{X} \) and \( \bar{Z} \) denote the population mean of study character \( y \), auxiliary character \( x \) and additional auxiliary character \( z \) having \( j \)th value \( Y_j, X_j \) and \( Z_j : j = 1,2,3,...,N \). Supposed the population of size \( N \) is divided in \( N_1 \) responding units and \( N_2 \) not responding unit. According to Hansen and Hurwitz a sample of size \( n \) is taken from population of size \( N \) by using simple random sampling without replacement (SRSWOR) scheme of sampling and it has been observed that \( n_1 \) units respond and \( n_2 \) units do not respond. Again by making extra effort, a sub sample of size \( r = n_2 k^{-1} \) is drawn from \( n_2 \) non-responding unit and collect information on \( r \) units for study character \( y \). Hence the estimator for \( \bar{Y} \) based on \( n_1 + r \) units on study character \( y \) is given by:

\[
\hat{Y} = \frac{n_1}{n} \bar{Y}_1 + \frac{n_2}{n} \bar{Y}_2
\]

where \( n_1 \) and \( n_2 \) are the responding and non-responding units in a sample of size \( n \) selected from population of size \( N \) by SRSWOR method of sampling. \( \bar{Y}_1 \) and \( \bar{Y}_2 \) are the means based on \( n_1 \) and \( r \) units selected from \( n_2 \) non-responding units by SRSWOR methods of sampling. Similarly we can also define estimator for population mean \( \bar{X} \) of auxiliary character \( x \) based on \( n_1 \) and \( r \) unit respectively, which is given as:

\[
\hat{X} = \frac{n_1}{n} \bar{X}_1 + \frac{n_2}{n} \bar{X}_2
\]

Variance of the estimators \( \hat{Y} \) and \( \hat{X} \) are given by

\[
V(\hat{Y}) = \frac{f}{n} S_y^2 + \frac{W_2(k-1)}{n} S_{y(2)}^2
\]

and

\[
V(\hat{X}) = \frac{f}{n} S_x^2 + \frac{W_2(k-1)}{n} S_{x(2)}^2
\]

where \( f = 1 - \frac{n}{N} \), \( W_2 = \frac{N_2}{N} \), \( (S_y^2, S_{y(2)}^2) \) and \( (S_x^2, S_{x(2)}^2) \) are population mean squares of \( y \) and \( x \) for entire population and non-responding part of population.

In case when the population means of the auxiliary character is unknown, we select a larger first phase sample of size \( n' \) units from a population of size \( N \) units by using simple random sample without replacement (SRSWOR) method of sampling and estimate \( \bar{X} \) by \( \bar{X}' \) based on these units \( n' \). Further second phase sample of size \( n \) (i.e. \( n < n' \)) is drawn from \( n' \) units by using SRSWOR method of sampling and variable \( y \) under investigation is measured \( n_1 \) responding and \( n_2 \) non-responding units. Again a sub sample of size \( r = \frac{n_2}{k} \) is drawn from \( n_2 \) non-responding units and collect information on \( r \) units by personal interview.
In this case two phase sampling ratio, product and regression estimators for population mean $\bar{Y}$ using one auxiliary character in the presence of non-response have been proposed by Khare and Srivastava (1993,1995) which are given as follows:

$$T_1 = \bar{y}^* \frac{\bar{x}^*}{\bar{x}^*}$$  \hspace{1cm} (5)

$$T_2 = \bar{y}^* + b^*(\bar{x}^* - \bar{x}^*)$$  \hspace{1cm} (6)

where $\bar{x}^*=\frac{n_1}{n}\bar{x}_1 + \frac{n_2}{n}\bar{x}_2$ , $\bar{x} = \frac{1}{n} \sum_{j=1}^{n} x_j$ , $\bar{x}' = \frac{1}{n'} \sum_{j=1}^{n'} x_j$ , $b^* = \frac{\hat{S}_{yx}}{\hat{S}_x^2}$ , $\hat{S}_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$

$\hat{S}_{yx}$ and $\hat{S}_x^2$ are estimates of $S_{yx}$ and $S_x^2$ based on $n_1 + r$ units.

The conventional and alternative two phase sampling ratio type estimators suggested by Khare and Srivastava (2010) which are as follows:

$$T_3 = \bar{y}^* \left( \frac{\bar{x}_x}{\bar{x}^*} \right)^\alpha$$  \hspace{1cm} (7)

$$T_4 = \bar{y}^* \left( \frac{\bar{x}_x}{\bar{x}} \right)^\alpha$$

where $\alpha$ and $\alpha'$ are constants.

Singh and Kumar (2010) have proposed difference type estimator using auxiliary character in the presence of non-response which is given as follows:

$$T_5 = \bar{y}^* \left( \frac{\bar{x}_x}{\bar{x}} \right)^{\alpha_1} \left( \frac{\bar{x}}{\bar{x}} \right)^{\alpha_2}$$  \hspace{1cm} (8)

where $\alpha_1$ and $\alpha_2$ are constants.

In case when $\bar{x}$ is not known than we may use an additional auxiliary character $z$ with known population mean $\bar{Z}$ with the assumption that the variable $z$ is less correlated to $y$ than $x$ i.e., $\rho_{yz} < \rho_{yx}$, $x$ and $z$ are variables such that $z$ is more cheaper than $x$. Following Chand (1975), some estimators have been proposed by Kiregyera (1980,84), Srivasatava et al. (1990) and Khare & Kumar (2011). In the case of non-response on the study character, the chain regression type and generalized chain type estimators for the population mean in the presence of non-response have been proposed by Khare & Kumar (2010) and Khare et al. (2011). An improved chain ratio-cum-regression type estimator for population mean in the presence of non-response have been proposed by Khare & Rehman (2014), which is given as follows:

$$T_6 = \bar{y}^* \left( \frac{\bar{x}_x}{\bar{x}} \right)^p \left( \frac{\bar{Z}}{\bar{Z}} \right)^q + b_{yx}(\bar{x} - \bar{x}') + b_{xz}(\bar{Z} - \bar{z}')$$  \hspace{1cm} (9)

where $p$ and $q$ are constants. $b_{yx}$ and $b_{xz}$ are regression coefficients. $\bar{Z}$ and $\bar{z}'$ population mean and sample mean based on first phase sample of size $n'$ units selected from population of size $N$ by SRSWOR method.
Mean Square Errors of the Study Estimator

Using the large sample approximations, the expressions for the mean square errors of the estimator proposed by Khare & Rehman (2014) up to the terms of order $n^{-1}$ are given by

$$MSE(T_k) = V(\bar{y}^*) + \left(1 - \frac{1}{n'}\right)\left\{ \bar{y}^2 p^2 C_x^2 + b_{yx}^2 X_x^2 - 2\bar{y}^2 p C_{yx} - 2\bar{y}b_{yx} C_{yx} + 2\bar{y} b_{yx} p C_{x}^2 \right\}$$

$$+ \left(1 - \frac{1}{n'}\right)\left\{ \bar{y}^2 q^2 C_z^2 + \bar{b}_z^2 Z_x^2 - 2\bar{y}^2 q C_{zx} - 2\bar{y} Z_b_{zx} C_{zx} + 2\bar{y} Z b_{zx} p C_{z}^2 \right\}$$

$$+ \frac{W_2 (K-1)}{n} \left\{ \bar{y}^2 p^2 C_{x(2)}^2 + b_{yx}^2 X_x^2 - 2\bar{y}^2 p C_{yx(2)} - 2\bar{y}b_{yx} C_{yx(2)} + 2\bar{y} b_{yx} p C_{x(2)}^2 \right\}$$

The optimum values of $p$ and $q$ and the values of regression coefficient are given as follows:

$$p_{opt} = \frac{\left(1 - \frac{1}{n'}\right)\bar{y}C_{yx} - \bar{X}b_{yx} C_z^2}{\bar{y}C_{z}^2}$$

$$q_{opt} = \frac{\bar{y}p_{yx} C_y}{\bar{Z} C_z}$$

and

$$b_{yx} = \frac{\bar{y}p_{yx} C_y}{\bar{X}}$$

Mean square errors of the estimators $T_1, T_2, T_3, T_4$ and $T_5$ are given as follows:

$$MSE(T_{min1}) = V(\bar{y}^*) + \bar{y}^2 \left(1 - \frac{1}{n'}\right)\left\{ C_x^2 - 2C_{yx} \right\} + \frac{W_2 (k-1)}{n} \left\{ C_{x(2)}^2 - 2C_{yx(2)} \right\}$$

$$MSE(T_{min2}) = V(\bar{y}^*) - \bar{y}^2 \left(1 - \frac{1}{n'}\right)\left\{ \rho_{yx}^2 C_y^2 + B^2 C_{x(2)}^2 - 2B C_{yx(2)} \right\}$$

$$MSE(T_{min3}) = V(\bar{y}^*) - \bar{y}^2 \left(1 - \frac{1}{n'}\right)\left\{ C_{yx}^2 + \frac{W_2 (k-1)}{n} C_{yx(2)} \right\}$$

$$MSE(T_{min4}) = V(\bar{y}^*) - \bar{y}^2 \left(1 - \frac{1}{n'}\right)\left\{ \rho_{yx}^2 C_y^2 \right\}$$

and

$$MSE(T_{min5}) = \bar{y}^2 \left[ \frac{f^2}{n'} C_y^2 + \left(1 - \frac{1}{n'}\right)\left(1 - \rho_{yx}^2 \right) C_y^2 + \frac{W_2 (k-1)}{n} \left(1 - \rho_{yx(2)}^2 \right) C_{y(2)}^2 \right]$$

where

$$V(\bar{y}^*) = \bar{y}^2 \left[ \frac{f^2}{n'} C_y^2 + \frac{W_2 (k-1)}{n} C_{y(2)}^2 \right]$$

and

$$B = \frac{\bar{y}p_{yx} C_y}{\bar{X} C_z}$$
**Determination of $n', n$ and $k$ for the Fixed Cost $C \leq C_0$**

Let us assume that $C_0$ be the total cost (fixed) of the survey apart from overhead cost. The expected total cost of the survey apart from overhead cost is given as follows:

$$C = (e'_1 + e'_2)n' + n\left( e_1 + e_2W_1 + e_3\frac{W_2}{k}\right),$$

(19)

where

- $e'_1$: the cost per unit of obtaining information on auxiliary character $x$ at the first phase.
- $e'_2$: the cost per unit of obtaining information on additional auxiliary character $z$ at the first phase.
- $e_1$: the cost per unit of mailing questionnaire/visiting the unit at the second phase.
- $e_2$: the cost per unit of collecting, processing data obtained from $n_1$ responding units.
- $e_3$: the cost per unit of obtaining and processing data (after extra efforts) for the sub sampling units.

The expression for $\text{MSE}(T_0)$ can be expressed in terms of $D_0, D_1, D_2$ and $D_3$ which are the coefficients $\frac{1}{n}, \frac{1}{n'}, \frac{1}{n}$ and $\frac{1}{N}$ respectively. The expression of $\text{MSE}(T_0)$ is given as follows:

$$\text{MSE}(T_0)_{\text{min}} = \frac{D_0}{n} + \frac{D_1}{n'} + \frac{kD_2}{n} - \frac{D_3}{N},$$

(20)

For obtaining the optimum values of $n', n, k$ for the fixed cost $C \leq C_0$, we define a function $\phi$ which is given as:

$$\phi = \text{MSE}(T_0)_{\text{min}} + \lambda(C - C_0),$$

(21)

where $\lambda$ is the Lagrange's multiplier.

We differentiating $\phi$ with respect to $n', n, k$ and equating zero, we get optimum values of $n', n$ and $k$ which are given as follows:

$$n'_{\text{opt}} = \sqrt{\frac{D_1}{\lambda(e'_1 + e'_2)}},$$

(22)

$$n_{\text{opt}} = \sqrt{\frac{(D_0 + k_{\text{opt}}D_2)}{\lambda\left(e_1 + e_2W_1 + e_3\frac{W_2}{k_{\text{opt}}}\right)}},$$

(23)

and

$$k_{\text{opt}} = \sqrt{\frac{D_0e_1W_2}{D_2(e_1 + e_2W_1)}},$$

(24)

where

$$\sqrt{\lambda} = \frac{1}{C_0}\left[\sqrt{D_1(e'_1 + e'_2)} + \sqrt{(D_0 + k_{\text{opt}}D_2)\left(e_1 + e_2W_1 + e_3\frac{W_2}{k_{\text{opt}}}\right)}\right],$$

(25)
The minimum value of \( MSE(T_0) \) for the optimum values of \( n', n \) and \( k \) in the expression \( MSE(T_0) \), we get:

\[
MSE(T_0)_{\text{min}} = \frac{1}{C_0} \left[ \sqrt{D_1(e_1' + e_2')} + \left( D_0 + k_{\text{opt}} D_2 \right) \left( e_1 + e_2 W_1 + e_3 \frac{W_2}{k_{\text{opt}}} \right) \right]^2 - \frac{D_2}{N},
\] (26)

Now neglecting the term of \( O(N^{-1}) \), we have

\[
MSE(T_0)_{\text{min}} = \frac{1}{C_0} \left[ \sqrt{D_1(e_1' + e_2')} + \left( D_0 + k_{\text{opt}} D_2 \right) \left( e_1 + e_2 W_1 + e_3 \frac{W_2}{k_{\text{opt}}} \right) \right]^2.
\] (27)

**Determination of \( n', n \) and \( k \) for the Specified Precision \( V = V_0' \)**

Let \( V_0' \) be the specified variance of the estimator \( T_0 \) which is fixed in advance, so we have

\[
V_0' = \frac{D_0}{n} + \frac{D_1}{n'} + \frac{k D_2}{n} - \frac{D_3}{N},
\] (28)

To find the optimum values of \( n', n, k \) and minimum expected total cost, we define a function \( \psi \) which is given as follows:

\[
\psi = (e_1' + e_2') n' + n \left( e_1 + e_2 W_1 + e_3 \frac{W_2}{k} \right) + \mu(MSE(T_0)_{\text{min}} - V_0') ,
\] (29)

where \( \mu \) is the Lagrange's multiplier.

After differentiating \( \psi \) with respect to \( n', n, k \) and equating to zero, we find the optimum values of \( n', n \) and \( k \) which are given as;

\[
n_{opt}' = \frac{\mu D_1}{(e_1' + e_2')},
\] (30)

\[
n_{opt} = \frac{\mu(D_0 + k_{\text{opt}} D_2)}{\left( e_1 + e_2 W_1 + e_3 \frac{W_2}{k_{\text{opt}}} \right)},
\] (31)

and

\[
k_{opt} = \frac{D_2 W_2 e_3}{D_2 (e_1 + e_2 W_1)},
\] (32)

where

\[
\sqrt{\mu} = \left[ \sqrt{D_1(e_1' + e_2')} + \left( D_0 + k_{\text{opt}} D_2 \right) \left( e_1 + e_2 W_1 + e_3 \frac{W_2}{k_{\text{opt}}} \right) \right]^2 - \frac{D_2}{N} + \frac{V_0' + D_3}{N},
\] (33)

The minimum expected total cost incurred on the use of \( T_0 \) for the specified variance \( V_0' \) will be given as follows:
Sampling Strategies for Finite Population Using Auxiliary Information

$$C_{\text{min}} = \frac{\sqrt{D_1(c'_1 + c'_2)} + \left[ (D_0 + k_{opt} D_2) \left( e_1 + e_2W_1 + e_3 \frac{W_2}{k_{opt}} \right) \right]^2}{V'_0 + \frac{D_0}{N}},$$

(34)

Now neglecting the terms of $O \left( \frac{1}{N^1} \right)$, we have

$$C_{\text{min}} = \frac{\sqrt{D_1(c'_1 + c'_2)} + \left[ (D_0 + k_{opt} D_2) \left( e_1 + e_2W_1 + e_3 \frac{W_2}{k_{opt}} \right) \right]^2}{V'_0},$$

(35)

An Empirical Study

To illustrate the results we use the data considered by Khare and Sinha (2007). The description of the population is given below:

The data on physical growth of upper socio-economic group of 95 schoolchildren of Varanasi under an ICMR study, Department of Pediatrics, B.H.U., during 1983-84 has been taken under study. The first 25% (i.e. 24 children) units have been considered as non-responding units. Here we have taken the study variable ($y$), auxiliary variable ($x$) and the additional auxiliary variable ($z$) are taken as follows:

$y$: weight (in kg.) of the children.
$x$: skull circumference (in cm) of the children.
$z$: chest circumference (in cm) of the children.

The values of the parameters of the $y$, $x$ and $z$ characters for the given data are given as follows:

$\bar{y} = 19.4968$, $\bar{z} = 51.1726$, $\bar{x} = 55.8611$, $C_y = 0.15613$, $C_z = 0.03006$, $C_x = 0.05860$, $C_{yz(1)} = 0.12075$, $C_{zx(1)} = 0.02478$, $C_{xy(2)} = 0.05402$, $\rho_{yz} = 0.328$, $\rho_{zx} = 0.846$, $\rho_{xy} = 0.297$,

$\rho_{xz(2)} = 0.570$, $W_2 = 0.25$, $W_1 = 0.74$, $N = 95$, $n = 35$

Table 1. Relative efficiency (in %) of the estimators with respect to $\hat{y}^*$ (for the fixed cost $C \leq C_0 = Rs. 220$, $c_1 = Rs. 0.90$, $c_2 = Rs. 0.10$, $c_1 = Rs. 2$, $c_2 = Rs. 4$, $c_3 = Rs. 25$).

<table>
<thead>
<tr>
<th>Estimators</th>
<th>$k_{opt}$</th>
<th>$n'_{opt}$</th>
<th>$n_{opt}$</th>
<th>Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{y}^*$</td>
<td>2.68</td>
<td>---</td>
<td>30</td>
<td>100 (0.3843)$^*$</td>
</tr>
<tr>
<td>$T_1$</td>
<td>2.89</td>
<td>58</td>
<td>23</td>
<td>117 (0.3272)</td>
</tr>
<tr>
<td>$T_2$</td>
<td>2.03</td>
<td>74</td>
<td>19</td>
<td>131 (0.2941)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>$T_3$</td>
<td>2.61</td>
<td>81</td>
<td>20</td>
<td>155 (0.2473)</td>
</tr>
<tr>
<td>$T_4$</td>
<td>1.06</td>
<td>76</td>
<td>14</td>
<td>136 (0.2819)</td>
</tr>
<tr>
<td>$T_5$</td>
<td>2.68</td>
<td>81</td>
<td>20</td>
<td>157 (0.2453)</td>
</tr>
<tr>
<td>$T_6$</td>
<td>2.67</td>
<td>68</td>
<td>21</td>
<td>166 (0.2315)</td>
</tr>
</tbody>
</table>

*Figures in parenthesis give the MSE (.).

From table 1, we obtained that for the fixed cost $C \leq C_0$ the study estimator $T_6$ is more efficient in comparison to the estimators $\bar{y}^*$, $T_1, T_2, T_3, T_4$ and $T_5$.

**Table 2.** Expected cost of the estimators for the specified variance $V_0' = 0.2356$ : ($c_1' = Rs. 0.90$, $c_2' = Rs. 0.10$, $c_1 = Rs. 2$, $c_2 = Rs. 5$, $c_3 = Rs. 25$)

<table>
<thead>
<tr>
<th>Estimators</th>
<th>$k_{opt}$</th>
<th>$n'_{opt}$</th>
<th>$n_{opt}$</th>
<th>Expected Cost (in Rs.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{y}^*$</td>
<td>2.68</td>
<td>---</td>
<td>61</td>
<td>502</td>
</tr>
<tr>
<td>$T_1$</td>
<td>2.89</td>
<td>107</td>
<td>40</td>
<td>418</td>
</tr>
<tr>
<td>$T_2$</td>
<td>2.03</td>
<td>115</td>
<td>25</td>
<td>332</td>
</tr>
<tr>
<td>$T_3$</td>
<td>2.61</td>
<td>88</td>
<td>20</td>
<td>246</td>
</tr>
<tr>
<td>$T_4$</td>
<td>1.06</td>
<td>92</td>
<td>16</td>
<td>275</td>
</tr>
<tr>
<td>$T_5$</td>
<td>2.68</td>
<td>87</td>
<td>21</td>
<td>244</td>
</tr>
<tr>
<td>$T_6$</td>
<td>2.67</td>
<td>69</td>
<td>20</td>
<td>231</td>
</tr>
</tbody>
</table>

From table 2, we obtained that for the specified variance the study estimator $T_6$ has less cost in comparison to the cost incurred in the estimators $\bar{y}^*$, $T_1, T_2, T_3, T_4$ and $T_5$. 
Conclusion

The information on additional auxiliary character and optimum values of increase the efficiency of the study estimators in comparison to corresponding estimators in case of the fixed cost $C \leq C_0$ and specified precision $V = V_0$.

References


The present book aims to present some improved estimators using auxiliary and attribute information in case of simple random sampling and stratified random sampling and in some cases when non-response is present.

This volume is a collection of five papers, written by seven co-authors (listed in the order of the papers): Sachin Malik, Rajesh Singh, Florentin Smarandache, B. B. Khare, P. S. Jha, Usha Srivastava and Habib Ur. Rehman.

The first and the second papers deal with the problem of estimating the finite population mean when some information on two auxiliary attributes are available. In the third paper, problems related to estimation of ratio and product of two population mean using auxiliary characters with special reference to non-response are discussed.

In the fourth paper, the use of coefficient of variation and shape parameters in each stratum, the problem of estimation of population mean has been considered. In the fifth paper, a study of improved chain ratio-cum-regression type estimator for population mean in the presence of non-response for fixed cost and specified precision has been made.

The authors hope that the book will be helpful for the researchers and students that are working in the field of sampling techniques.