Rajesh Singh ■ Florentin Smarandache (editors)

SAMPLING STRATEGIES FOR FINITE POPULATION USING AUXILIARY INFORMATION

The Educational Publisher
Columbus, 2015

Rajesh Singh ■ Florentin Smarandache (editors)

SAMPLING STRATEGIES FOR FINITE POPULATION
USING AUXILIARY INFORMATION

Papers by Sachin Malik, Rajesh Singh, Florentin Smarandache,
B. B. Khare, P. S. Jha, Usha Srivastava, Habib Ur. Rehman.

The Educational Publisher
Zip Publishing
1313 Chesapeake Ave.
Columbus, Ohio 43212, USA
Email: info@edupublisher.com

ISBN 978-1-59973-348-7

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

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

[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:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$s_{\phi_1\phi_2}=\frac{1}{n-1}\sum_{i=1}^{n}\left(\phi_{1i}-p_1\right)\left(\phi_{2i}-p_2\right)$ and $\rho_\phi=\frac{s_{\phi_1\phi_2}}{s_{\phi_1}s_{\phi_2}}$ be the sample phi-covariance and phi-correlation between ϕ_1 and ϕ_2 respectively, corresponding to the population phi-covariance and phi-correlation $S_{\phi_1\phi_2}=\frac{1}{N-1}\sum_{i=1}^{N}\left(\phi_{1i}-P_1\right)\left(\phi_{2i}-P_2\right)$

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 β_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

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

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

$$MSE(t_5)=\overline{Y}^2f_1\left[C_y^2+C_{p_1}^2\left(\alpha_1^2-2\alpha_1K_{pb_1}\right)+C_{p_2}^2\left(\alpha_2^2-2\alpha_2K_{pb_2}+2\alpha_1\alpha_2K_{\phi}\right)\right] \tag{2.17}$$

$$MSE(t_6)=\overline{Y}^2f_1\left[C_y^2+C_{p_1}^2\left(\frac{\beta_1^2}{4}-\beta_1K_{pb_1}\right)+C_{p_2}^2\left(\frac{\beta_2^2}{4}-\frac{\beta_1\beta_2}{2}K_{\varphi}+\beta_2K_{pb_1}\right)\right] \tag{2.18}$$

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

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

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

Where,

$$t_0=\overline{y}\,,\ t_1=\overline{y}\left[\frac{L_1P_1+L_2}{L_1p_1+L_2}\right]^{\alpha_1}\left[\frac{L_3P_2+L_4}{L_3p_2+L_4}\right]^{\alpha_2}$$

and $t_2=\exp\left[\frac{(L_5P_1+L_6)-(Lp_1+L_6)}{(L_1P_1+L_2)+(L_5p_1+L_6)}\right]^{\beta_1}\exp\left[\frac{(L_7p_2+L_6)-(L_7P_2+L_8)}{(L_7p_2+L_2)+(L_7P_2+L_8)}\right]^{\beta_2}$

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

$$t_p=\overline{Y}\left(1+e_0\right)\left[\begin{array}{l}w_0+w_1\left(1+\varphi_1e_1\right)^{-\alpha_1}\left(1+\varphi_2e_2\right)^{-\alpha_2}\\+w_2\exp\left(-\theta_1e_1\left[1+\theta_1e_1\right]^{-1}\right)^{\beta_1}\\\exp\left(-\theta_2e_2\left[1+\theta_2e_2\right]^{-1}\right)^{\beta_2}\end{array}\right] \tag{3.3}$$

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

$$MSE\left(t_p\right) = \overline{Y}^2 f\left[w_1^2 T_1 + w_2^2 T_2 + 2w_1 w_2 T_3 - 2w_1 T_4 - 2w_2 T_5\right] \tag{3.5}$$

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

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

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,

ρ_{pb_1}=0.621, ρ_{pb_2}=0.673, ρ_ϕ=0.889.

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

CHOICE OF SCALERS, when $w_0 = 0\ w_1 = 1\ w_2 = 0$						
α_1	α_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}	ρ_{pb_1}	C_{p_2}	ρ_{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	ρ_{pb_1}	N	ρ_{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, $w_0 = 0\ w_1 = 0\ w_2 = 1$						

β_1	β_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}	ρ_{pb_1}	C_{p_2}	ρ_{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	ρ_{pb_1}	N	ρ_{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 ϕ_1 is unknown but the same is known for the second auxiliary attribute ϕ_2. When P_1 is unknown, it is some times estimated from a preliminary large sample of size n' on which only the attribute ϕ_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

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

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

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

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

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

$$B(t_{d1}) = \bar{Y}f_3C_{p_1}^2\left[1 - k_{pb_1}\right] \tag{5.5}$$

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

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

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

$$MSE(t_{d1}) = \bar{Y}^2\left[f_1C_y^2 + f_3C_{P_1}^2\left(1 - 2K_{pb_1}\right)\right] \tag{5.9}$$

$$MSE(t_{d2}) = \bar{Y}^2\left[f_1C_y^2 + f_2C_{p_2}^2\left(1 - 2K_{kp_2}\right)\right] \tag{5.10}$$

$$MSE(t_{d3}) = \bar{Y}^2\left[f_1C_y^2 + f_3\frac{C_{p_1}^2}{4}\left(1 - 4K_{pb_1}\right)\right] \tag{5.11}$$

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

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

$$B(t_{d5}) = \bar{Y}\left[f_3C_{p_1}^2\left(\frac{m_1^2}{2} + \frac{m_1}{2} - m_1K_{pb_1}\right) + f_2C_{P_2}^2\left(\frac{m_2^2}{2} + \frac{m_2}{2} - m_2k_{pb_2}\right)\right] \tag{5.15}$$

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

$$MSE(t_{d5}) = \bar{Y}\left[f_1C_y^2 + f_3C_{p_1}^2\left(m_1^2 - 2m_1K_{pb_1}\right) + f_2C_{p_2 2}^2\left(m_2^2 - 2m_2K_{pb_2}\right)\right] \tag{5.17}$$

$$MSE(t_{d6}) = \bar{Y}^2\left[f_1C_y^2 + f_3\left(\frac{n_1^2}{4} - n_1K_{pb_1}\right)C_{p_1}^2 + f_2\left(\frac{n_2^2}{4} + n_2K_{pb_2}\right)C_{p_2}^2\right] \tag{5.18}$$

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,

$$t_0 = \bar{y}\,,\; t_{d1} = \bar{y}\left[\frac{L_1p'_1 + L_2}{L_1p_1 + L_2}\right]^{m_1}\left[\frac{L_3P_2 + L_4}{L_3p'_2 + L_4}\right]^{m_2}$$

$$\text{and } t_{d2} = \exp\left[\frac{(L_5p'_1 + L_6) - (Lp_1 + L_6)}{(L_1p'_1 + L_2) + (L_5p_1 + L_6)}\right]^{n_1}\exp\left[\frac{(L_7p'_2 + L_6) - (L_7P_2 + L_8)}{(L_7p'_2 + L_2) + (L_7P_2 + L_8)}\right]^{n_2}$$

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

$$t_p = \overline{Y}\left(1+e_0\right)\left[h_0 + h_1\left(1+\varphi_1 e'_1\right)^{m_1}\left(1+\varphi_1 e_1\right)^{-m_1}\left(1+\varphi_2 e'_2\right)^{-m_2}\right.$$

$$\left. + h_2 \exp\left(\theta_1\left[e'_1 - e_1\right]\left[1+\theta_1\left(e'_1 - e_1\right)\right]^{-1}\right)^{n_1}\exp\left(\theta_2 e'_2\left[1+\theta_2 e'_2\right]\right)^{n_2}\right] \tag{6.3}$$

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

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

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

$$MSE\left(t_{pd}\right) = \overline{Y}^2\left[h_1^2 R_1 + h_2^2 R_2 + 2h_1h_2R_3 + 2h_1R_4 + 2h_2R_5\right] \tag{6.5}$$

where,

$$\left.\begin{aligned} h_1 &= \frac{R_2R_4 - R_3R_5}{R_1R_2 - R_3^2} \\ h_2 &= \frac{R_1R_5 - R_3R_4}{R_1R_2 - R_3^2} \end{aligned}\right\} \tag{6.6}$$

and

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

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, $\overline{Y}$=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,

ρ_{pb_1}=0599, ρ_{pb_2}=0.559, ρ_{ϕ}=0.725.

Table 6.1: PRE of different estimators of $\overline{Y}$ 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}	ρ_{pb_1}	C_{p_2}	ρ_{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	ρ_{pb_1}	N	ρ_{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}	ρ_{pb_1}	C_{p_2}	ρ_{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	ρ_{pb_1}	N	ρ_{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.

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

[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 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 ϕ_i denote the observations on variable y and ϕ respectively for i^{th} unit ($i = 1, 2, ..., N$).

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 ϕ. Let P=A/N and p=a/n denote the proportion of units in the population and sample respectively possessing attribute ϕ. 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

$$t_{NG} = \bar{y}\frac{P}{p} \tag{1.1}$$

with MSE

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

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

$$S_\phi^2 = \frac{1}{N-1}\sum_{i=1}^{N}(\phi_i - P)^2, \quad S_{y\phi} = \frac{1}{N-1}\sum_{i=1}^{N}(\phi_i - P)\left(y_i - \bar{Y}\right).$$

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

$$q = p + mA = p + NmP \text{ , and}$$

$$Q = (Nm+1)P,$$

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

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

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

where w_1, w_2 and α are suitably chosen scalars.

To obtain the Bias and MSE of the estimator t, we write

$$\bar{y} = \bar{Y}(1+e_0)\ ,\ p = P(1+e_1)\ ,\ s_\phi^2 = S_\phi^2(1+e_2),$$

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

such that $E(e_i) = 0$, i=0,1,2,3 and

$$E(e_0^2) = \left(\frac{1}{n}-\frac{1}{N}\right)C_y^2\ , \qquad E(e_1^2) = \left(\frac{1}{n}-\frac{1}{N}\right)C_p^2\ ,$$

$$E(e_0e_1) = \left(\frac{1}{n}-\frac{1}{N}\right)\rho_{pb}C_yC_p, \quad E(e_1e_2) = \left(\frac{1}{n}-\frac{1}{N}\right)C_p\lambda_{03},$$

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

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

$$t = \bar{Y}\left[w_1(1+e_0) - w_2\frac{\beta}{R}e_1(1+e_3)(1+e_2)^{-1}\right]\left(1+\frac{e_1}{Nm+1}\right)^\alpha \tag{2.2}$$

We assume that $|e_2|<1$ and $\left|\frac{e_1}{Na+1}\right|<1$, so that $(1+e_2)^{-1}$ 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

$$t-\bar{Y} = \bar{Y}[w_1\left\{1+e_0+\frac{\alpha e_1}{Nm+1}+\frac{\alpha(\alpha-1)}{2}\frac{e_1^2}{(Nm+1)^2}+\frac{\alpha e_0e_1}{Nm+1}\right\}$$

$$-w_2\frac{\beta}{R}\left\{e_1+e_1e_3-e_1e_2+\frac{\alpha e_1^2}{Nm+1}\right\}-1] \tag{2.3}$$

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

$$B(t) = \bar{Y}\left[(w_1-1)+w_1\left\{\frac{\alpha}{Nm+1}f_1C_{py}+\frac{\alpha(\alpha-1)}{2(Nm+1)^2}f_1C_p^2\right\}\right.$$

$$\left.-w_2\frac{\beta}{R}f_1\left\{C_p\frac{\lambda_{12}}{\rho_{pb}}-C_p\lambda_{03}+\frac{\alpha}{Nm+1}C_p^2\right\}\right] \tag{2.4}$$

Squaring both sides of (2.3) and neglecting terms of e's having power greater than two, we have

$$\left(t-\overline{Y}\right)^2=\overline{Y}^2\left[w_1^2\left\{1+2e_0+\frac{2\alpha e_1}{Nm+1}+e_0^2+\frac{\alpha(2\alpha-1)e_1^2}{(Nm+1)^2}+\frac{4\alpha e_0e_1}{Nm+1}\right\}\right.$$
$$+w_2^2\left(\frac{\beta}{R}\right)^2e_1^2+1-2w_1w_2\frac{\beta}{R}\left\{e_1+e_1e_3+e_0e_1-e_1e_2+\frac{2\alpha e_1^2}{Nm+1}\right\}$$
$$-2w_1\left\{1+e_0+\frac{\alpha e_1}{Nm+1}+\frac{\alpha e_0e_1}{Nm+1}+\frac{\alpha(\alpha-1)e_1^2}{2(Nm+1)^2}\right\}$$
$$+2w_2\frac{\beta}{R}\left\{e_1+e_1e_3-e_1e_2+\frac{\alpha e_1^2}{Nm+1}\right\} \tag{2.5}$$

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

$$MSE(t)=\overline{Y}^2\left[1+w_1^2A_{1(\alpha)}^{(m)}+w_2^2A_2-2w_1w_2A_{3(\alpha)}^{(m)}-2w_1A_{4(\alpha)}^{(m)}+2w_2A_{5(\alpha)}^{(m)}\right] \tag{2.6}$$

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

$$A_2=\left(\frac{\beta}{R}\right)^2f_1C_p^2$$

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

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

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

where , $k=\rho_{pb}\dfrac{C_y}{C_p}$.

The MSE(t) is minimised for

$$w_1=\frac{\left(A_2A_{4(\alpha)}^{(m)}-A_{3(\alpha)}^{(m)}A_{5(\alpha)}^{(m)}\right)}{\left(A_{1(\alpha)}^{(m)}A_2-\left(A_{3(\alpha)}^{(m)}\right)^2\right)}=w_{10} \tag{2.7}$$

$$w_2=\frac{\left(A_{3(\alpha)}^{(m)}A_{4(\alpha)}^{(m)}-A_{1(\alpha)}^{(m)}A_{5(\alpha)}^{(m)}\right)}{\left(A_{1(\alpha)}^{(m)}A_2-\left(A_{3(\alpha)}^{(m)}\right)^2\right)}=w_{20} \tag{2.8}$$

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

Choice of scalars				Estimator
w_1	w_2	α	m	
1	0	0	0	$t_1 = \bar{y}$
w_1	0	0	0	$t_2 = w_1\bar{y}$ Searls (1964) type estimator
w_1	0	α	m	$t_3 = w_1\bar{y}\left(\frac{q}{Q}\right)^{\alpha}$
w_1	0	α	0	$t_4 = w\ \bar{y}\left(\frac{p}{P}\right)^{\alpha}$
1	0	-1	0	$t_5 = \bar{y}\left(\frac{P}{p}\right)$, Naik and Gupta (1996) estimator
1	1	-1	0	$t_6 = \left[\bar{y} + b(P-p)\right]\frac{P}{p}$ Singh et. al. (2008) estimator
w_1	w_2	0	0	$t_7 = \left[w_1\bar{y} + w_2 b(P-p)\right]$
w_1	1	0	0	$t_8 = \left[w_1\bar{y} + b(P-p)\right]$
w	w	0	0	$t_9 = w\left[\bar{y} + b(P-p)\right]$
1	1	0	0	$t_{10} = \bar{y} + b(P-p)$ Regression estimator

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

$$Var(t_1) = f_1 S_y^2 \tag{3.1}$$

To, the first degree of approximation the biases and MSE's of t_i's, i=1,2,.......,10 are respectively given by

$$B(t_2)=\overline{Y}(w_1-1) \tag{3.2}$$

$$B(t_3)=\overline{Y}\left[(w_1-1)+f_1w_1\left\{\frac{\alpha}{Nm+1}\rho_{pb}C_yC_p+\frac{\alpha(\alpha-1)}{2(Nm+1)}C_p^2\right\}\right] \tag{3.3}$$

$$B(t_4)=\overline{Y}\left[(w_1-1)+w_1f_1\left\{\alpha\rho_{pb}C_yC_p+\frac{\alpha(\alpha-1)}{2}C_p^2\right\}\right] \tag{3.4}$$

$$B(t_5)=\overline{Y}f_1\left[C_p^2-\rho_{pb}C_yC_p\right] \tag{3.5}$$

$$B(t_6)=\overline{Y}f_1\left[(C_p^2-\rho_{pb}C_yC_p)-\frac{\beta}{R}\left(C_p\frac{\lambda_{12}}{\rho_{pb}}-C_p\lambda_{03}\right)\right] \tag{3.6}$$

$$B(t_7)=\overline{Y}\left[(w_1-1)-\frac{\lambda_{12}\beta}{R}f_1\left\{C_p\frac{\lambda_{12}}{\rho_{pb}}-C_p\lambda_{03}\right\}\right] \tag{3.7}$$

$$B(t_8)=\overline{Y}\left[(w_1-1)-\frac{\beta}{R}f_1\left\{C_p\frac{\lambda_{12}}{\rho_{pb}}-C_p\lambda_{03}\right\}\right] \tag{3.8}$$

$$B(t_9)=\overline{Y}\left[(w-1)-wf_1\frac{\beta}{R}\left\{C_p\frac{\lambda_{12}}{\rho_{pb}}-C_p\lambda_{03}\right\}\right] \tag{3.9}$$

$$B(t_{10})=-\overline{Y}\frac{\beta}{R}f_1\left[C_p\frac{\lambda_{12}}{\rho_{pb}}-C_p\lambda_{03}\right] \tag{3.10}$$

The corresponding MSE's will be

$$MSE(t_2)=\overline{Y}^2\left[1+w_1^2A_{1(0)}^{(0)}-2w_1A_{4(0)}^{(0)}\right] \tag{3.11}$$

$$MSE(t_3)=\overline{Y}^2\left[1+w_1^2A_{1(\alpha)}^{(m)}-2w_1A_{4(\alpha)}^{(m)}\right] \tag{3.12}$$

$$MSE(t_4)=\overline{Y}^2\left[1+w_1^2A_{1(\alpha)}^{(0)}-2w_1A_{4(\alpha)}^{(0)}\right] \tag{3.13}$$

$$MSE(t_5)=\overline{Y}^2\left[1+A_{1(-1)}^{(0)}-2A_{4(-1)}^{(0)}\right] \tag{3.14}$$

$$MSE(t_6) = \overline{Y}^2\left[1 + A^{(0)}_{1(-1)} + A_2 - 2A^{(0)}_{3(-1)} - 2A^{(0)}_{4(-1)} + 2A^{(0)}_{5(-1)}\right] \tag{3.15}$$

$$MSE(t_7) = \overline{Y}^2\left[1 + w_1^2 A^{(0)}_{1(0)} + w_2^2 A_2 - 2w_1 w_2 A^{(0)}_{3(0)} - 2w_1 A^{(0)}_{4(0)} + 2w_2 A^{(0)}_{5(0)}\right] \tag{3.16}$$

$$MSE(t_8) = \overline{Y}^2\left[1 + w_1^2 A^{(0)}_{1(0)} + A_2 - 2w_1\left(A^{(0)}_{3(0)} + A^{(0)}_{4(0)}\right) + 2A^{(0)}_{5(0)}\right] \tag{3.17}$$

$$MSE(t_9) = \overline{Y}^2\left[1 + w^2\left(A^{(0)}_{1(0)} + A_2 - 2A^{(0)}_{3(0)}\right) - 2w\left(A^{(0)}_{4(0)} - A^{(0)}_{5(0)}\right)\right] \tag{3.18}$$

$$MSE(t_{10}) = \overline{Y}^2\left[1 + A^{(0)}_{1(0)} + A_2 - 2\left(A^{(0)}_{3(0)} + A^{(0)}_{4(0)} - A^{(0)}_{5(0)}\right)\right] \tag{3.19}$$

The MSE's of the estimaors of t_i, i=2,3,4,7,8,9 will be minimised respectively, for

$$w_1 = \frac{A^{(0)}_{4(0)}}{A^{(0)}_{1(0)}} \tag{3.20}$$

$$w_1 = \frac{A^{(m)}_{4(\alpha)}}{A^{(m)}_{1(\alpha)}} \tag{3.21}$$

$$w_1 = \frac{A^{(0)}_{4(\alpha)}}{A^{(0)}_{1(\alpha)}} \tag{3.22}$$

$$\left.\begin{aligned} w_1 &= \frac{\left(A_2 A^{(0)}_{4(0)} - A^{(0)}_{3(0)} A^{(0)}_{5(0)}\right)}{\left(A_2 A^{(0)}_{1(0)} - \left(A^{(0)}_{3(0)}\right)^2\right)} \\ w_2 &= \frac{\left(A^{(0)}_{3(0)} A^{(0)}_{4(0)} - A^{(0)}_{1(0)} A^{(0)}_{5(0)}\right)}{A_2 A^{(0)}_{1(0)} - \left(A^{(0)}_{3(0)}\right)^2} \end{aligned}\right\} \tag{3.23}$$

$$w_1 = \frac{\left(A^{(0)}_{3(0)} + A^{(0)}_{4(0)}\right)}{A^{(0)}_{1(0)}} \tag{3.24}$$

$$w = \frac{A^{(0)}_{4(0)} - A^{(0)}_{5(0)}}{A^{(0)}_{1(0)} + A_2 - 2A^{(0)}_{3(0)}} \tag{3.25}$$

Thus the resulting minimum MSE of t_i , i= 2,3,4,7,8,9 are, respectively given by

$$\min.\ MSE(t_2) = \overline{Y}^2\left[1 - \frac{\left(A_{4(0)}^{(0)}\right)^2}{A_{1(0)}^{(0)}}\right] \tag{3.26}$$

$$\min.\ MSE(t_3) = \overline{Y}^2\left[1 - \frac{\left(A_{4(\alpha)}^{(m)}\right)^2}{A_{1(\alpha)}^{(m)}}\right] \tag{3.27}$$

$$\min.\ MSE(t_4) = \overline{Y}^2\left[1 - \frac{\left(A_{4(\alpha)}^{(0)}\right)^2}{A_{1(\alpha)}^{(0)}}\right] \tag{3.28}$$

$$\min.\ MSE(t_7) = \overline{Y}^2\left[1 - \frac{\left\{A_2\left(A_{4(0)}^{(0)}\right)^2 - 2A_{3(0)}^{(0)}A_{5(0)}^{(0)} + A_{1(0)}^{(0)}\left(A_{5(0)}^{(0)}\right)^2\right\}}{\left\{A_2A_{1(0)}^{(0)} - A_{3(0)}^{(0)}\right\}^2}\right] \tag{3.29}$$

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

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

4. *Empirical study*

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

y = Number of villages in the circles and

ϕ = 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 $\overline{y}$.

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

Estimator	PRE's
$t_1 = \bar{y}$	100.00
t_2	101.41
t_3	90.35
t_4	6.92
t_5	11.64
t_6	7.38
t_7	100.44
t_8	243.39
t_9	243.42
t_{10}	241.98

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(t10), 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

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. Bedi, P. K. (1996). Efficient utilization of auxiliary information at estimation stage. Biom. Jour, 38:973–976.

3. 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.

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

5. Searls, D.T. (1964): The utilization of known coefficient of variation in the estimation procedure. Journ. of the Amer. Stat. Assoc., 59, 1125-1126.

6. 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.

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. Sukhatme, P.V. and Sukhatme, B.V. (1970): Sampling theory of surveys with applications. Iowa State University Press, Ames, U.S.A.

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:

$$R^* = f(v,u) \text{ and } P^* = g(w,u) \tag{1}$$

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 $u=\left(\frac{\bar{x}_1'}{\bar{x}_1}\right)$. 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:

$$R_g = \hat{R}h(u',v') \tag{2}$$

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:

$$R_1^* = \hat{R}\left(\frac{\bar{y}_3}{\bar{y}_3'}\right)\left(\frac{\bar{y}_4'}{\bar{Y}_4}\right) \quad \text{and} \quad R_2^* = \hat{R}\left(\frac{\bar{y}_3'}{\bar{y}_3}\right)\left(\frac{\bar{Y}_4}{\bar{y}_4'}\right) \tag{3}$$

$$P_1^* = \hat{P}\left(\frac{\bar{y}_3}{\bar{y}_3'}\right)^{\alpha_1}\left(\frac{\bar{y}_4'}{\bar{Y}_4}\right)^{\alpha_2} \quad \text{and} \quad P_2^* = \hat{P}\left(\frac{\bar{y}_3}{\bar{Y}_3}\right)^{\beta_1}\left(\frac{\bar{y}_4'}{\bar{Y}_4}\right)^{\beta_2} \tag{4}$$

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

$$\hat{R}_h = h(\hat{R},u,v) \text{and}\ \hat{P}_h = h(\hat{P},u,v), \tag{5}$$

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:

$$R^* = \hat{R}h(u_1,u_2..u_p) = \hat{R}h(\underline{u}) \text{ and } R^{**} = g(\hat{R},\underline{u}), \tag{6}$$

such that $h(\underline{e})=1$ and $g(\hat{R},\underline{e})=R$, where $h(\underline{u})$ and $g(\hat{R},\underline{u})$ 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:

$$R_m^* = g(\hat{R},\underline{u'}) , \tag{7}$$

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

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:

$$\theta = \frac{\bar{Y}_1, \bar{Y}_2, \bar{Y}_3, ..., \bar{Y}_m}{\bar{Y}_{m+1}, \bar{Y}_{m+2}, \bar{Y}_{m+3}, ..., \bar{Y}_k} , \tag{8}$$

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

$$\hat{\theta} = \frac{\bar{y}_1, \bar{y}_2, \bar{y}_3, ..., \bar{y}_m}{\bar{y}_{m+1}, \bar{y}_{m+2}, \bar{y}_{m+3}, ..., \bar{y}_k} , \tag{9}$$

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

$m = 2, k = 2; \quad \theta = P$

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

$m = k = 1; \bar{Y}_1 = \bar{Y}_2 , \quad \theta = \bar{Y}_1^{\,2} ,$

$m = k = 3; \bar{Y}_1 = \bar{Y}_2 = \bar{Y}_3 , \quad \theta = \bar{Y}_1^{\,3} ,$

$m = 2, k = 4; \bar{Y}_1 = \bar{Y}_2, \bar{Y}_3 = \bar{Y}_4 , \quad \theta = \bar{Y}_1^{\,2} = R^2 ,$

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

$$\theta^* = \hat{\theta} h(\underline{u}), \tag{10}$$

such that $h(\underline{e})=1$, where $\underline{u}'=(u_1,u_2,...u_p)$ and $u_i=\dfrac{\bar{x}_i}{\bar{X}_i}$, $i=1, 2, ..., p$.

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

a) Whatever be the sample chosen $(\underline{u})$, assume values in abounded closed convex sub set G of p dimensional real space containing the point $\underline{u}=\underline{e}$.

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

c) The first and second partial derivatives of $h(\underline{u})$ 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 θ^* 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 θ^* 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:

$$\theta^{**}=\hat{\theta}h(\underline{v}), \tag{11}$$

where $\underline{v}=(v_1,v_2,...v_p)'$, $v_i=\dfrac{\bar{x}_i'}{\bar{x}_i}$, $i=1,2,...p$.

Such that $h(\underline{e})=1$ and $h(\underline{v})$ 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:

$$R_i^*=R^*h(u_i)\text{and } P_i^*=P^*h(u_i), i=1,2, \tag{12}$$

such that $h(1)=1$, where $R^*=\dfrac{\bar{y}_1^*}{\bar{y}_2^*}$, $P^*=\bar{y}_1^*\bar{y}_2^*$, $u_1=\dfrac{\bar{x}^*}{\bar{X}}$, $u_2=\dfrac{\bar{x}}{\bar{X}}$ 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:

$$R_i = g(v,u_i)\ i = 1,2, \tag{13}$$

such that $g(R,1) = R$, $g_1(R,1)$ and $g_{12}(R,1) = R^{-1} g_2(R,1)$, where $v = \frac{\bar{y}_1^*}{\bar{y}_2^*}$, $u_1 = \bar{x}^*$ and $u_2 = \bar{x}$. The function $g(v,u_i)\ i = 1,2$ assumes positive values in a real line containing the point $(R,1)$. 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 $g(v,u_i)\ i = 1,2$ 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 $g(v,u_i)\ i = 1,2$ with respect to v, v and u_i, 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:

$$C_1 = w_0 v u_i^{\alpha_i},\ C_2 = w_1 v + w_2 u_i,\ C_3 = w_1' v + w_2' v u_i^{\beta_i},\ i = 1,2, \tag{14}$$

where w_0, w_1, w_2, w_1', w_2', α_i and β_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:

$$T_i = g(v,u_i),\ i = 1,2, \tag{15}$$

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

$$T_i^* = g(w,u_i),\ i = 1,2, \tag{16}$$

such that $g(P,1) = 1$ and $g_1(P,1) = 1$, where $w = \bar{y}_1^* \bar{y}_2^*$, $u_1 = \frac{\bar{x}^*}{\bar{x}'}$, $u_2 = \frac{\bar{x}}{\bar{x}'}$ 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:

$$T_{g1} = \hat{R}\left(\frac{\bar{x}^*}{\bar{x}'}\right)^{\alpha_1}\left(\frac{\bar{z}'}{\bar{Z}}\right)^{\alpha 2} \text{ and } T_{g2} = \hat{R}\left(\frac{\bar{x}}{\bar{x}'}\right)^{\beta_1}\left(\frac{\bar{z}'}{\bar{Z}}\right)^{\beta 2}, \tag{17}$$

where $\hat{R} = \frac{\bar{y}_1^*}{\bar{y}_2^*}$ and (α_1, α_2) and (β_1, β_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 $R^* = k_1 \hat{R}$ and $P^* = k_1 \hat{P}$ 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:

$$R_{ci} = f\left(\hat{R}, u_i, v\right),\ i = 1,2, \tag{18}$$

such that $f(R,1,1) = 1$ and $f_1(R,1,1) = 1$, where $\hat{R} = \dfrac{\bar{y}_1^*}{\bar{y}_2^*}$, $u_1 = \dfrac{\bar{x}^*}{\bar{x}'}$, $u_2 = \dfrac{\bar{x}}{\bar{x}'}$ and $v = \dfrac{\bar{z}'}{\bar{Z}}$. The function $f\left(\hat{R}, u_i, v\right)$, $i = 1,2$ 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:

$$t_i = \hat{R} g_i(\underline{u}_i'),\ i = 1, 2, \tag{19}$$

such that $g_i(\underline{e}_i') = 1$, where $\underline{u}_i$ and $\underline{e}_i$ denote the column vectors $(u_{i1}, u_{i2}, \ldots, u_{ip})'$ and $(1, 1, \ldots, 1)'$, $u_{1j} = \dfrac{\bar{x}_j^*}{\bar{X}_j}$ and $u_{2j} = \dfrac{\bar{x}_j}{\bar{X}_j}$ $j = 1, 2, \ldots, p$.

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

1. Hansen, M. H. and Hurwitz, W. N. (1946): The problem of non-response in sample surveys. Jour. Amer. Stat. Assoc., 41, 517-529.

2. Khare, B. B. (1987): On modified class of estimators of ratio and product of two population means using auxiliary character. Proc. Math. Soc, B.H.U. 3, 131-137.

3. Khare, B. B. (1990): A generalized class of estimators for combination of products and ratio of some population means using multi-auxiliary characters. J. Stat. Res., 24, 1-8.

4. Khare, B. B. (1991 (a)): Determination of sample sizes for a class of two phase sampling estimators for ratio and product of two population means using auxiliary character. Metron (Italy), XLIX, (1-4), 185-197.

5. Khare, B. B. (1991 (b)): On generalized class of estimators for ratio of two population means using multi-auxiliary characters. Aligarh J. Stat. 11, 81-90.

6. Khare, B. B. (1992): On class of estimators for product of two population means using multi-auxiliary characters with known and unknown means. Ind. J. Appl. Stat., 1, 56-67.

7. Khare, B. B. (1993(a)): A class of two phase sampling estimators for the combination of product and ratio of several population means using multi-auxiliary characters. Proc. Nat. Acad. Sci., India, 63 (4), Pt. II, 391-397.

8. Khare, B. B. (1993(b)): On a class of two phase sampling estimators for ratio of two population means using multi-auxiliary characters. Proc. Nat. Acad. Sci., India, 63(a), III, 513-520.

9. Khare, B. B. (2014): Estimation of population parameters using the technique of sub sampling from non respondents in sample surveys- A Review. Proc. Nat. Acad. Sci. Sec A, 84 (3), 337-343.

10. Khare, B. B. (2014-15): Applications of statistics in bio-medical sciences. Prajna, Special Issue on Science & Technology, Vol. - 60 (2),.

11. Khare, B. B. and Pandey, S. K. (2000): A class of estimators for ratio of two population means using auxiliary character in presence of non-response. J. Sc. Res., 50, 115-124.

12. Khare, B. B. and Sinha, R. R. (2002a): Estimation of the ratio of two populations means using auxiliary character with unknown population mean in presence of non response. Prog. Maths. Vol.- 36. No. (1, 2), 337-348.

13. Khare, B. B. and Sinha, R. R. (2002b): On class of two phase sampling estimators for the product of two population means using auxiliary character in presence of non response. Proc. of Vth international symposium on optimization and Statistics held at AMU, Aligarah, 221-232.

14. Khare, B. B. and Sinha, R. R. (2004 (a)): Estimation of finite population ratio using two phase sampling in presence of non response. Aligarh J. Stat. 24, 43-56.

15. Khare, B. B. and Sinha, R. R. (2004 (b)): On the general class of two phase sampling estimators for the product of two population means using the auxiliary characters in the presence of non-response. Ind.J. Appl. Statistics., 8, 1-14.

16. Khare, B. B. and Sinha, R. R. (2007): Estimation of the ratio of the two population means using multi- auxiliary characters in presence of non-response. In "Statistical techniques in life testing, reliability, sampling theory and quality control" edited by B. N. Pandey Narosa publishing house, New Delhi, 163-171.

17. Khare, B. B. and Sinha, R. R. (2010): On class of estimators for the product of two population means using auxiliary character in presence of non-response. Inter. Trans. Appl. Sci., 2(4), 841-846.

18. Khare, B. B. and Sinha, R. R. (2012 (a)): Combined class of estimators for ratio and product of two population means in presence of non-response. Int. Jour. Stats. and Eco., 8(S12), 86-95.

19. Khare, B. B. and Sinha, R. R. (2012 (b)): Improved classes of ratio of two population means with double sampling the non-respondents. Statistika-Statistics & Economy Jour. 49(3), 75-83.

20. Khare, B. B. and Sinha, R. R. (2014): A class of two phase sampling estimator for ratio of two populations means using multi-auxiliary characters in the presence of non response. Stat. in Trans. New series, 15, (3), 389-402.

21. Khare, B. B. and Srivastava, S. R. (1999): A class of estimators for ratio of two population means and means of two populations using auxiliary character. J. Nat. Acad. Math. 13, 100-104.

22. Khare, B. B. and Srivastava, S. Rani. (1998): Combined generalized chain estimators for ratio and product of two population means using auxiliary characters. Metron (Italy) LVI (3-4), 109-116.

23. Khare, B. B., Jha, P. S. and Kumar, K. (2014): Improved generalized chain estimators for ratio and product of two population means using two auxiliary characters in the presence of non-response. International J. Stats & Economics, 13(1), 108-121.

24. Khare, B. B., Pandey, S. K. and Kumar, A. (2013 (a)): Improved class of estimators for ratio of two population means using auxiliary character in presence of non-response. Proc. Nat. Acad. Sci. India, 83(1), 33-38.

25. Khare, B. B., Kumar, K. and Srivastava, U. (2013 (b)): Improved classes of chain type estimators for ratio of two population means using two auxiliary characters in the presence of non-response. Int. Jour. Adv. Stats. & Prob. 1 (3), 53-63.

26. Khare, B. B., Srivastava, U. and Kumar K. (2012): Chain type estimators for ratio of two population means using auxiliary characters in the presence of non response. J. Sc. Res., BHU, 56, 183-196.

27. Khare, B. B., Srivastava, U. and Kumar K. (2013): Generalized chain type estimators for ratio of two population means using two auxiliary characters in the presence of non-response. International J. Stats & Economics, 10(1), 51-64.

28. Rao, J. N. K. and Pareira, N. P. (1968): On double ratio estimators. Sankhya Ser. A., 30, 83-90.

29. Ray, S. K. and Singh, R. K. (1985): Some estimators for the ratio and product of population parameters. Jour. Ind. Soc. Agri. Stat., 37, 1-10.

30. Shah, S. M. and Shah, D. N. (1978): Ratio cum product estimators for estimating ratio (product) of two population parameters. Sankhya Ser. *C.,* 40, 156-166.

31. Singh, M. P. (1969): Comparison of some ratio cum product estimators. Sankhya, Ser. B, 31, 375-378.

32. Singh, R. K. (1982b): On estimating ratio and product of population parameters. Calcutta Stat. Assoc., 20, 39-49.

33. Singh, V. K., Singh, Hari P. and Singh, Housila P. (1994): Estimation of ratio and product of two finite population means in two phase sampling. J. Stat. Plan. Inf. 41, 163-171.

34. Singh, V. K., Singh, Hari P., Singh, Housila P. and Shukla, D. (1994): A general class of chain type estimators for ratio and product of two population means of a finite population. Commun. Stat.- TM. 23 (5), 1341-1355.

35. Srivastava, Rani S., Khare, B. B. and Srivastava, S.R. (1988): On generalized chain estimator for ratio and product of two population means using auxiliary characters. Assam Stat. Review, 2(1), 21-29.

36. Srivastava, Rani S., Srivastava, S. R. and Khare, B. B. (1989): Chain ratio type estimator for ratio of two population means using auxiliary characters. Commun. Stat. Theory Math.(USA), 18(10), 3917-3926.

37. Tripathi, T. P. (1980): A general class of estimators for population ratio. Sankhya Ser. C., 42, 63-75.

38. Tripathi, T. P., Das, A. K. and Khare, B. B. (1994): Use of auxiliary information in sample surveys - A review. Aligarh J. Stat., 14, 79-134.

On The Use of Coefficient of Variation and β_1, β_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 $\beta_{1i}, \beta_{2i}\ (i = 1,2,...K)$ 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 ith stratum and n_i denotes the size of the sample to be selected from the ith stratum and h be the number of strata with

$$\sum_{i=1}^{h} N_i = N \text{ and } \sum_{i=1}^{h} n_i = n \ , \tag{1}$$

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 $\overline{\text{Y}}_{\text{N}}$ can be expressed as

$$\overline{\text{Y}}_{\text{N}} = \sum_{i=1}^{h} p_i \overline{Y}_i \ , \tag{2}$$

where $p_i = \dfrac{N_i}{N}$ 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 $\bar{y}_i$ and s_i^2 respectively. Then the estimate of $\overline{Y}_N$ is given by

$$\text{T}_1 = \sum_{i=1}^{h} p_i \bar{y}_i \tag{3}$$

and the

$$\text{V}(\text{T}_1) = \sum_{i=1}^{h} \frac{p_i^2 \sigma_i^2}{n_i} \quad \text{(if f.p.c is ignored)}, \tag{4}$$

where σ_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

$$\text{T}_2 = \sum_{i=1}^{h} p_i \{\alpha_i \bar{y}_i + (1-\alpha_i) C_i^{-1} \sqrt{s_i^2}\,\} \tag{5}$$

and expectation of T_2 is given by

$$\begin{aligned}
\text{E}(\text{T}_2) &= \sum_{i=1}^{h} p_i \{\alpha_i \overline{Y}_i + (1-\alpha_i)\overline{Y}_i (1 - \frac{1}{8}\frac{V(s_i^2)}{\sigma_i^4})\} \\
&= \sum_{i=1}^{h} p_i \{\overline{Y}_i - (1-\alpha_i)(\frac{1}{8}\frac{V(s_i^2)}{\sigma_i^4})\} \\
&= \sum_{i=1}^{h} p_i \{\overline{Y}_i - \frac{(1-\alpha_i)}{8}(\frac{\beta_{2i}-1}{n_i} + \frac{2}{n(n-1)})\} \\
&= \overline{Y}_N + O(\frac{1}{n_i})
\end{aligned}$$

(6)

where β_{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

$$\mathrm{MSE}(\mathrm{T}_2/\alpha_i) = \sum_{i=1}^{h} \frac{p_i^2 \sigma_i^2}{n_i} \left[\{\alpha_i^2 + \alpha_i(1-\alpha_i)C_i^{-1}\sqrt{\beta_{1i}} + \frac{(1-\alpha_i)^2}{4} C_i^{-2}(\beta_{2i}-1)\} + O(\frac{1}{n^{3/2}}) \right]. \quad (7)$$

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

$$\alpha_{\mathrm{i}opt} = \frac{\beta_{2i} - 2C_i\sqrt{\beta_{1i}} - 1}{4C_i^2 + 4C_i\sqrt{\beta_{1i}} + \beta_{2i} - 1}, \quad (8)$$

where β_{1i} 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

$$\mathrm{MSE}(\mathrm{T}_2)_{\min} = \sum_{i=1}^{h} \frac{p_i^2 \sigma_i^2}{n_i} \left[\{\frac{\beta_{2i} - \beta_{1i} - 1}{(\beta_{2i} - \beta_{1i} - 1) + (\sqrt{\beta_{1i}} - 2C_i)^2}\} + O(\frac{1}{n^{3/2}}) \right]. \quad (9)$$

The value of $\alpha_{i\,opt}$ will be less than one for $\sqrt{\beta_{1i}} < 2C_i$, 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 $\sqrt{\beta_{1i}} = 2C_i$, which is true for gamma and exponential distribution. The value of $\alpha_{i\,opt}$ will be greater than one for $\sqrt{\beta_{1i}} > 2C_i$, 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 $\sqrt{\beta_{1i}} < 2C_i$ or $\sqrt{\beta_{1i}} > 2C_i$, 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 $\sqrt{\beta_{1i}} = 2C_i$ 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, β_{1i}'s and β_{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

$$\mathrm{T}_3 = \sum_{i=1}^{h} p_i \{\alpha_i' \bar{y}_i + (1-\alpha_i')c_i^{-1}\sqrt{s_i^2}\} \tag{10}$$

The mean square error of the estimator T_3 as given by

$$\mathrm{MSE}(\mathrm{T}_2/\alpha_i') = \sum_{i=1}^{h} \frac{p_i^2\sigma_i^2}{n_i}\left[\alpha_i'^2 + \alpha_i'(1-\alpha_i')C_i^{-1}\sqrt{\beta_{1i}} + (1-\alpha_i')^2 C_i^{-2}\{(\beta_{2i}-1) + 4n_i C_i^{-2}V(c_i)\}\right], \tag{11}$$

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

The optimum value of α_i' is given by

$$\alpha_{1opt}' = \frac{(\beta_{2i}-\beta_{1i}-1) + 4n_i C_i^{-2}V(c_i)}{(\beta_{2i}-\beta_{1i}-1+4n_i C_i^{-2}V(c_i)) + (\sqrt{\beta_{1i}}-2C_i)^2}\ . \tag{12}$$

It is easy to see that

$$\mathrm{MSE}(\mathrm{T}_3/\alpha_{1opt}')_{\min} = \sum_{i=1}^{h} \frac{p_i^2\sigma_i^2}{n_i}\left[\frac{(\beta_{2i}-\beta_{1i}-1) + 4n_i C_i^{-2}V(c_i)}{(\beta_{2i}-\beta_{1i}-1+4n_i C_i^{-2}V(c_i)) + (\sqrt{\beta_{1i}}-2C_i)^2}\right]. \tag{13}$$

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

Case 3: C_i's, β_{1i}'s and β_{2i}'s are unknown:

When C_i's , β_{1i}'s and β_{2i}'s are not known then they can be estimated on the basis of a larger sample of size $n_i' >> ...n_i$ from the past data and we may have the estimator for the population mean $\bar{Y}_N$ given by

$$\mathrm{T}_4 = \sum_{i=1}^{h} p_i \{\hat{\alpha}_i \bar{y}_i + (1-\hat{\alpha}_i)c_i^{-1}\sqrt{s_i^2}\}\ , \tag{14}$$

where $\hat{\alpha}_{1opt} = \dfrac{\hat{\beta}_{2i} - 2\hat{C}_i\sqrt{\hat{\beta}_{1i}} - 1}{4\hat{C}_i^2 + 4\hat{C}_i\sqrt{\hat{\beta}_{1i}} + \hat{\beta}_{2i} - 1}$.

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

References

1. Searls, D. T. (1964): The utilization of coefficient of variation in the estimation procedure. Jour. of Amer. Stat. Assoc., 59, 1125-1126.

2. Searls, D. T. (1967): A note on the use of a approximately known coefficient of variation. The Amer. Statistician, 21, 20-21.

3. Sen, A. R. and Gerig, T. M. (1975): Estimation of a population mean having equal coefficient of variation on succession occasions. Bull. Int. Stat. Inst., 46, 314-22.

4. Sen, A. R. (1978): Estimation of the population mean when the coefficient of variation is known. Commu. Stat. Theory Meth., A7, 1, 657-672.

5. Srivastava, S.K. and Jhajj, H.S. (1983): A class of estimators of the population means using multi- auxiliary information. Calcutta Stat. Assoc. Bull, 32, 47-56.

6. Searls, D. T. and Intarapanich R (1990): A note on an estimator for variance that utilized the kurtosis. Amer. Stat., 44(4), 295-296.

7. Hirano K (1972): Using some approximately known coefficient of variation in estimating mean. Proc. Inst. Stat. Math, 20(2), 61-64.

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

[1]B. B. Khare , [1]Habib Ur Rehman 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, 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 jth 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:

$$\bar{y}^* = \frac{n_1}{n}\bar{y}_1 + \frac{n_2}{n}\bar{y}_2' \tag{1}$$

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;

$$\bar{x}^* = \frac{n_1}{n}\bar{x}_1 + \frac{n_2}{n}\bar{x}_2' \tag{2}$$

Variance of the estimators $\bar{y}^*$ and $\bar{x}^*$ are given by

$$V(\bar{y}^*) = \frac{f}{n}S_y^2 + \frac{W_2(k-1)}{n}S_{y(2)}^2 \tag{3}$$

and

$$V(\bar{x}^*) = \frac{f}{n}S_x^2 + \frac{W_2(k-1)}{n}S_{x(2)}^2 \tag{4}$$

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 ($n_2 / k, k > 1$) 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:

$$T_1 = \bar{y}^* \frac{\bar{x}'}{\bar{x}^*} \tag{5}$$

$$T_2 = \bar{y}^* + b^*\left(\bar{x}' - \bar{x}^*\right) \tag{6}$$

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

$\hat{S}_{yx}$ 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:

$$T_3 = \bar{y}^*\left(\frac{\bar{x}^*}{\bar{x}'}\right)^{\alpha}, \qquad T_4 = \bar{y}^*\left(\frac{\bar{x}}{\bar{x}'}\right)^{\alpha'} \tag{7}$$

where
α and α' 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:

$$T_5 = \bar{y}^*\left(\frac{\bar{x}}{\bar{x}^*}\right)^{\alpha_1}\left(\frac{\bar{x}'}{\bar{x}}\right)^{\alpha_2} \tag{8}$$

where α_1 and α_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:

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

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

$$MSE(T_6)=V\left(\bar{y}^*\right)+\left(\frac{1}{n}-\frac{1}{n'}\right)\left\{\bar{Y}^2p^2C_x^2+b_{yx}^2\bar{X}^2C_x^2-2\bar{Y}^2pC_{yx}-2\bar{X}\bar{Y}b_{yx}C_{yx}+2\bar{X}\bar{Y}b_{yx}pC_x^2\right\}$$
$$+\left(\frac{1}{n'}-\frac{1}{N}\right)\left\{\bar{Y}^2q^2C_z^2+b_{xz}^2\bar{Z}^2C_z^2-2\bar{Y}^2qC_{yz}-2\bar{Y}\bar{Z}b_{xz}C_{yz}+2\bar{Y}\bar{Z}qb_{xz}C_z^2\right\}$$
$$+\frac{W_2(K-1)}{n}\left\{\bar{Y}^2p^2C_{x(2)}^2+b_{yx}^2\bar{X}^2C_{x(2)}^2-2\bar{Y}^2pC_{yx(2)}-2\bar{X}\bar{Y}b_{yx}C_{yx(2)}+2\bar{X}\bar{Y}b_{yx}pC_{x(2)}^2\right\} \quad (10)$$

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

$$p_{opt}=\frac{\left(\frac{1}{n}-\frac{1}{n'}\right)\left(\bar{Y}C_{yx}-\bar{X}b_{yx}C_x^2\right)+\frac{W_2(k-1)}{n}\left(\bar{Y}C_{yx(2)}-\bar{X}b_{yx}C_{x(2)}^2\right)}{\left(\frac{1}{n}-\frac{1}{n'}\right)\bar{Y}C_x^2+\frac{W_2(k-1)}{n}\bar{Y}C_{x(2)}^2} \quad (11)$$

$$q_{opt}=\frac{\bar{Y}C_{yz}-\bar{Z}b_{xz}C_z^2}{\bar{Y}C_z^2}, \quad (12)$$

$$b_{yx}=\frac{\bar{Y}\rho_{yx}}{\bar{X}}\frac{C_y}{C_x} \quad \text{and} \quad b_{yx}=\frac{\bar{X}\rho_{xz}}{\bar{Z}}\frac{C_x}{C_z} \quad (13)$$

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

$$MSE(T_1)_{\min}=V(\bar{y}^*)+\bar{Y}^2\left[\left(\frac{1}{n}-\frac{1}{n'}\right)\left(C_x^2-2C_{yx}\right)+\frac{W_2(k-1)}{n}\left(C_{x(2)}^2-2C_{yx(2)}\right)\right] \quad (14)$$

$$MSE(T_2)_{\min}=V(\bar{y}^*)-\bar{Y}^2\left[\left(\frac{1}{n}-\frac{1}{n'}\right)\rho_{yx}^2C_y^2+\frac{W_2(k-1)}{n}\left\{B^2C_{x(2)}^2-2B\,C_{yx(2)}\right\}\right] \quad (15)$$

$$MSE(T_3)_{\min}=V(\bar{y}^*)-\bar{Y}^2\left[\frac{\left\{\left(\frac{1}{n}-\frac{1}{n'}\right)C_{yx}+\frac{W_2(k-1)}{n}C_{yx(2)}\right\}^2}{\left(\frac{1}{n}-\frac{1}{n'}\right)C_x^2+\frac{W_2(k-1)}{n}C_{x(2)}^2}\right] \quad (16)$$

$$MSE(T_4)_{\min}=V(\bar{y}^*)-\bar{Y}^2\left(\frac{1}{n}-\frac{1}{n'}\right)\rho_{yx}^2C_y^2 \quad (17)$$

and

$$MSE(T_5)_{\min}=\bar{Y}^2\left[\frac{f'}{n'}C_y^2+\left(\frac{1}{n}-\frac{1}{n'}\right)(1-\rho_{yx}^2)C_y^2+\frac{W_2(k-1)}{n}(1-\rho_{yx(2)}^2)C_{y(2)}^2\right] \quad (18)$$

where $V(\bar{y}^*)=\bar{Y}^2\left\{\frac{f}{n}C_y^2+\frac{W_2(k-1)}{n}C_{y(2)}^2\right\}$ and $B=\frac{\bar{Y}\rho_{yx}C_y}{\bar{X}C_x}$

Determination of n', n and k *for the Fixed Cost* $C \le C_0$

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:

$$C = (e_1' + e_2')n' + n\left(e_1 + e_2 W_1 + e_3 \frac{W_2}{k}\right), \tag{19}$$

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, $MSE(T_6)$ can be expressed in terms of D_0, D_1, D_2 and D_3 which are the coefficients $\frac{1}{n}$, $\frac{1}{n'}$, $\frac{k}{n}$ and $\frac{1}{N}$ respectively. The expression of $MSE(T_6)$ is given as follows:

$$MSE(T_6)_{\min} = \frac{D_0}{n} + \frac{D_1}{n'} + \frac{k\,D_2}{n} - \frac{D_3}{N} \quad , \tag{20}$$

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

$$\phi = MSE(T_6)_{\min} + \lambda(C - C_0), \tag{21}$$

where λ is the Lagrange's multiplier.

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

$$n'_{opt} = \sqrt{\frac{D_1}{\lambda(e_1' + e_2')}} \quad , \tag{22}$$

$$n_{opt} = \sqrt{\frac{(D_0 + k_{opt} D_2)}{\lambda\left(e_1 + e_2 W_1 + e_3 \frac{W_2}{k_{opt}}\right)}} \quad , \tag{23}$$

and

$$k_{opt} = \sqrt{\frac{D_0 e_3 W_2}{D_2(e_1 + e_2 W_1)}} \quad , \tag{24}$$

where

$$\sqrt{\lambda} = \frac{1}{C_0}\left[\sqrt{D_1(e_1' + e_2')} + \sqrt{(D_0 + k_{opt} D_2)\left(e_1 + e_2 W_1 + e_3 \frac{W_2}{k_{opt}}\right)}\right] \quad , \tag{25}$$

The minimum value of $MSE(T_6)$ for the optimum values of n', n and k in the expression $MSE(T_6)$, we get:

$$MSE(T_6)_{\min} = \frac{1}{C_0}\left[\sqrt{D_1(e_1'+e_2')} + \sqrt{(D_0 + k_{opt}D_2)\left(e_1 + e_2W_1 + e_3\frac{W_2}{k_{opt}}\right)}\right]^2 - \frac{D_3}{N} \ , \tag{26}$$

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

$$MSE(T_6)_{\min} = \frac{1}{C_0}\left[\sqrt{D_1(e_1'+e_2')} + \sqrt{(D_0 + k_{opt}D_2)\left(e_1 + e_2W_1 + e_3\frac{W_2}{k_{opt}}\right)}\right]^2 \tag{27}$$

Determination of n', n and k *for the Specified Precision* $V = V_0'$

Let V_0' be the specified variance of the estimator T_6 which is fixed in advance, so we have

$$V_0' = \frac{D_0}{n} + \frac{D_1}{n'} + \frac{kD_2}{n} - \frac{D_3}{N} \ , \tag{28}$$

To find the optimum values of n', n, k and minimum expected total cost, we define a function ψ which is give as follows:

$$\psi = (e_1'+e_2')n' + n\left(e_1 + e_2W_1 + e_3\frac{W_2}{k}\right) + \mu(MSE(T_6)_{\min} - V_0') \ , \tag{29}$$

where μ is the Lagrange's multiplier.

After differentiating ψ with respect to n', n, k and equating to zero, we find the optimum value of n', n and k which are given as;

$$n'_{opt} = \sqrt{\frac{\mu D_1}{(e_1'+e_2')}} \ , \tag{30}$$

$$n_{opt} = \sqrt{\frac{\mu(D_0 + k_{opt}D_2)}{\left(e_1 + e_2W_1 + e_3\frac{W_2}{k_{opt}}\right)}} \ , \tag{31}$$

and

$$k_{opt} = \sqrt{\frac{D_0W_2e_3}{D_2(e_1 + e_2W_1)}} \ , \tag{32}$$

where

$$\sqrt{\mu} = \frac{\left[\sqrt{D_1(e_1'+e_2')} + \sqrt{(D_0 + k_{opt}D_2)\left(e_1 + e_2W_1 + e_3\frac{W_2}{k_{opt}}\right)}\right]^2}{V_0' + \frac{D_3}{N}} \ , \tag{33}$$

The minimum expected total cost incurred on the use of T_6 for the specified variance V_0' will be given as follows:

$$C_{6\min} = \frac{\left[\sqrt{D_1(e_1' + e_2')} + \sqrt{(D_0 + k_{opt}D_2)\left(e_1 + e_2W_1 + e_3\dfrac{W_2}{k_{opt}}\right)}\right]^2}{V_0' + \dfrac{D_3}{N}}, \tag{34}$$

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

$$C_{6\min} = \frac{\left[\sqrt{D_1(e_1' + e_2')} + \sqrt{(D_0 + k_{opt}D_2)\left(e_1 + e_2W_1 + e_3\dfrac{W_2}{k_{opt}}\right)}\right]^2}{V_0'}, \tag{35}$$

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:

$\bar{Y} = 19.4968$, $\bar{Z} = 51.1726$, $\bar{X} = 55.8611$, $C_y = 0.15613$, $C_z = 0.03006$, $C_x = .05860$, $C_{y(2)} = 0.12075$, $C_{z(2)} = 0.02478$, $C_{x(2)} = 0.05402$, $\rho_{yz} = 0.328$, $\rho_{yx} = 0.846$, $\rho_{xz} = 0.297$, $\rho_{xz(2)} = 0.570$, $W_2 = 0.25$, $W_1 = 0.74, N = 95, n = 35$

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

Estimators	k_{opt}	n'_{opt}	n_{opt}	Efficiency
$\bar{y}^*$	2.68	---	30	100 (0.3843)*
T_1	2.89	58	23	117 (0.3272)
T_2	2.03	74	19	131 (0.2941)

T_3	2.61	81	20	155 (0.2473)
T_4	1.06	76	14	136 (0.2819)
T_5	2.68	81	20	157 (0.2453)
T_6	2.67	68	21	166 (0.2315)

*Figures in parenthesis give the MSE (.).

From table 1, we obtained that for the fixed cost $C \le C_0$ 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*)

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

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 $C \le C_0$ and specified precision $V = V_0$.

References

1. Chand, L. (1975): Some ratio-type estimators based on two or more auxiliary variables. Ph.D. Thesis submitted to Iowa State University, Ames, IOWA.

2. Hansen, M. H. and Hurwitz, W. N. (1946): The problem of non-response in sample surveys. *J. Amer. Statist. Assoc.,* 41, 517-529.

3. Kiregyera, B. (1980): A chain ratio type estimator in finite population double sampling using two auxiliary variables. *Metrika,* 27, 217-223.

4. Kiregyera, B. (1984): Regression - type estimators using two auxiliary variables and model of double sampling from finite populations. *Metrika,* 31, 215-226.

5. Khare, B. B. and Srivastava, S. (1993): Estimation of population mean using auxiliary character in presence of non-response. *Nat. Acad. Sci. Letters, India*, 16(3), 111-114.

6. Khare, B. B. and Srivastava, S. (1995): Study of conventional and alternative two phase sampling ratio, product and regression estimators in presence of non-response. *Proc. Nat. Acad. Sci., India,* Sect. A 65(a) II, 195-203.

7. Khare, B. B. and Srivastava, S. (1996): Transformed product type estimators for population mean in presence of softcore observations. *Proc. Math. Soc. B. H.U.,* 12, 29-34.

8. Khare, B. B. and Srivastava, S. (1997): Transformed ratio type estimators for population mean in presence. *Commun. Statist. Theory Meth.,* 26(7), 1779-1791.

9. Khare, B.B. and Sinha, R.R. (2007): Estimation of the ratio of the two populations means using multi-auxiliary characters in the presence of non-response. In “Statistical Technique in Life Testing, Reliability, Sampling Theory and Quality Control”. Edited by B.N. Pandey, Narosa publishing house, New Delhi, 163-171.

10. Khare, B.B., Kumar Anupam, Sinha R.R., Pandey S.K.(2008): Two phase sampling estimators for population mean using auxiliary character in presence of non-response in sample surveys. *Jour. Scie. Res., BHU, Varanasi*, 52: 271-281.

11. Khare, B.B. and Kumar, S. (2009): Transformed two phase sampling ratio and product type estimators for population mean in the presence of nonresponse. *Aligarh J. Stats.,* 29, 91-106.

12. Khare, B.B. and Srivastava, S. (2010): Generalized two phase sampling estimators for the population mean in the presence of nonresponse. *Aligarh. J. Stats.,* 30, 39-54.

13. Khare, B. B. and Kumar, S. (2010): Chain regression type estimators using additional auxiliary variable in two phase sampling in the presence of non response. *Nat. Acad. Sci. Letters, India*, 33, No. (11 & 12), 369-375.

14. Khare, B.B. and Kumar, S. (2011): A generalized chain ratio type estimator for population mean using coefficient of variations of the study variable. *Nat. Acad. Sci. Letters, India,* 34(9-10), 353-358.

15. Khare, B.B., Srivastava, U. and Kamlesh Kumar (2011): Generalized chain estimators for the population mean in the presence of non-response. *Proc. Nat. Acad. Sci., India,* 81(A), pt III. 231-238.

16. Khare, B.B., and Rehman, H. U. (2014): An improved chain ratio-cum-regression type estimator for population mean in the presence of non-response. Int. J. Agri. Stat. Sci., 10(2), 281-284.

17. Rao,P.S.R.S.(1986):Ratio estimation with sub-sampling the non-respondents. *Survey Methodology*. 12(2):217-230.

18. Rao, P.S.R.S.(1990): Ratio and regression Estimators with Sub-sampling of the non-respondents.In: Liepine, Guner E.,Uppuluri VRR, editors. Data Quality Control theory and Pragmatics*,* Marcel Dekker, New York. pp.191-208.

19. Singh, H.P. and Kumar, S. (2010): Estimation of mean in presence of non-response using two phase sampling scheme. *Statistical papers,* 51,559-582.

20. Srivastava, S. R., Khare, B.B. and Srivastava, S.R .(1990): A generalised chain ratio estimator for mean of finite population. *Jour. Ind. Soc. Agri. Stat.,* 42(1), 108-117.

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.

ISBN 978-1-59973-348-7

9 781599 733487 >