STATISTICA, anno LXXIII, n. 2 2013

THREE PHASE STRATIFIED SAMPLING WITH RATIO METHOD OF ESTIMATION

Bijoy Kumar Pradhan

Former Faculty Member, Department of Statistics, Utkal University, Bhubaneswar, Odisha, India

1. INTRODUCTION

Sometimes a survey sampler selects a large sample of units to collect information on certain variables and then select a relatively smaller sample to collect information on main character under study. This is the problem of two phase sampling. Further phases may be added, if required and the resulting sampling is termed as multiphase sampling, when the sample for the main survey is selected in more than two phases. For example, let us consider the problem of estimating the total consumer expenditure in a particular city through a sample survey, when the available information is only a list of all households in the city. If it is decided to select a sample of households, this might involve a very large sample to get a reasonably precise estimate and hence the cost involved may be considerable and prohibitive. An alternative procedure may be thought of as selecting a preliminary moderately large sample in the first phase to collect information on characteristics such as household size, occupational status etc. and these information may be used either for stratification, or for selection or in estimation procedures. In the second phase, a sub-sample from the first phase sample or an independent sample from the population is selected to observe the main character under study.

Although two-phase sampling can theoretically be extended to three or more phases, not much theoretical work is available concerning such extensions. In this paper we treat a three-phase sampling where the first phase is used for stratification, the second phase sample to observe auxiliary variables to estimate the auxiliary population characteristics to be used in estimation and the third phase sample to observe the main character under study.

The 'double sampling' technique was first formulated by Neyman (1938) in connection with the collection of information on strata sizes in a stratified sampling experiment. Further contributions are due to Robson (1952), Robson and King (1953), Pascual (1961), Rao (1973) and many others.

2. NOTATIONS

Assume that the population U of size N is divided into L strata. Further for the h^{th} stratum $\left(h=1\ ,\ 2\ ,\ldots,\ L\right)$ denote

N_h the number of units for the h^{th} stratum

$n_h^{'}$ the number of units falling into h^{th} stratum s_{1h} after stratifying the first phase sample $s_1\left(n^{'}\right)$ with the help of stratifying variable Z , $\sum_{h=1}^{L} n_h^{'} = n^{'}$.

$n_h^{''}$ the number of units in a subsample s_{2h} from $s_{1h}\left(n_h^{'}\right)$ for each h in the second phase to observe an auxiliary variable x .

$n_h^{'''}$ the number of units in a subsample s_{3h} drawn from $s_{2h}\left(n_h^{''}\right)$ for each h in the third phase to observe the main character y .

$W_h = \dfrac{N_h}{N}$, $w_h = \dfrac{n_h^{'}}{n^{'}}$

$g_h = \dfrac{n_h^{''}}{n_h^{'}}$, a constant proportion of units sampled from the h^{th} stratum at the second phase, $0 < g_h \leq 1$.

$t_h = \dfrac{n_h^{'''}}{n_h^{''}}$, a constant proportion of unit sampled from the h^{th} stratum at the third phase, $0 < t_h \leq 1$.

$\bar{Y}_h = \dfrac{1}{N_h}\sum_{i=1}^{N_h} y_{hi}$, the mean based on N_h units of y .

$\bar{X}_h = \dfrac{1}{N_h}\sum_{i=1}^{N_h} x_{hi}$, the mean based on N_h units of x .

$\bar{y}_h^{'}$ the sample mean based on first phase sample of size in the h^{th} stratum of y .

$\bar{y}_h^{''}$ the sample mean based on second phase sample of size $n_h^{''}$ in the h^{th} stratum of y .

$\bar{y}_h^{'''}$ the sample mean based on third phase sample of size $n_h^{'''}$ in the h^{th} stratum of y .

$\bar{x}_h^{'}$ the sample mean based on first phase sample of size $n_h^{'}$ in the h^{th} stratum of x .

$\bar{x}_h^{''}$ the sample mean based on second phase sample of size $n_h^{''}$ in the h^{th} stratum of x .

$\bar{x}_h'''$ the sample mean based on third phase sample of size n_h''' in the h^{th} stratum of x.

$R_h = \dfrac{\bar{Y}_h}{\bar{X}_h}$, $R_h' = \dfrac{\bar{y}_h'}{\bar{x}_h'}$, $R_h'' = \dfrac{\bar{y}_h''}{\bar{x}_h''}$

$S_{yh}^2 = \dfrac{1}{N_h - 1}\sum_{i=1}^{N_h}\left(y_{hi} - \bar{Y}_h\right)^2$, the mean squared error based on N_h units of y.

$S_{xh}^2 = \dfrac{1}{N_h - 1}\sum_{i=1}^{N_h}\left(x_{hi} - \bar{X}_h\right)^2$, the mean squared error based on N_h units of x.

S_{yxh} population covariance between y and x in the h^{th} stratum

$S_{rh}^2 = S_{yh}^2 + R_h^2 S_{xh}^2 - 2R_h S_{yxh}$

$s_{yh}'^2$ the sample mean squared error based on first phase sample of y in the h^{th} stratum

$s_{xh}'^2$ the sample mean squared error based on first phase sample of x in the h^{th} stratum

$s_{yxh}'^2$ the sample covariance based on first phase sample of y and x in the h^{th} stratum

$s_{rh}'^2 = s_{yh}'^2 + R_h'^2 s_{xh}'^2 - 2R_h' s_{yxh}'$

$s_{yh}''^2$ the sample mean squared error based on second phase sample of y in the h^{th} stratum

$s_{xh}''^2$ the sample mean squared error based on second phase sample of x in the h^{th} stratum

$s_{yxh}''^2$ the sample covariance of y and x based on second phase sample in the h^{th} stratum

ρ_h the population correlation coefficient between x and y for the h^{th} stratum.

3. DOUBLE SAMPLING FOR STRATIFICATION

Considering double sampling for stratification

$$\bar{\bar{y}}_{st} = \sum_{h=1}^{L} w_h \bar{y}_h \tag{3.1}$$

where preliminary simple random sample without replacement (SRSWOR) sample s_1 of n' size is drawn from the whole population to estimate W_h and the second sample which is a subsample of first where $n_h^{''}$ units are selected according to simple random sample without replacement (SRSWOR) out of $n_h^{'}$ units belonging to stratum h in the initial sample, under the assumption that an unbiased estimate of W_h is $w_h = n_h^{'}/n^{'}$, $\overline{\overline{y}}_{st}$ is an unbiased estimator of $\overline{Y}$ and the large sample variance of $\overline{\overline{y}}_{st}$ is given by

$$V\left(\overline{\overline{y}}_{st}\right) = \left(\frac{1}{n^{'}} - \frac{1}{N}\right) S_y^2 + \sum_{h=1}^{L} \left(\frac{1}{g_h} - 1\right) \frac{W_h S_{yh}^2}{n^{'}} \tag{3.2}$$

Minimizing the $V\left(\overline{\overline{y}}_{st}\right)$ subject to the cost function

$$C = c_1 n^{'} + \sum_{h=1}^{L} c_h n_h^{''} , \tag{3.3}$$

where c_1 is the cost of observing a unit in the first phase and c_h is the cost of observing a second phase unit, the optimum value of variance is obtained as

$$V\left(\overline{\overline{y}}_{st}\right)_{opt} = \frac{\left(\sqrt{\left(S_y^2 - \sum W_h S_{yh}^2\right) c_1} + \sum W_h S_{yh} \sqrt{c_h}\right)^2}{C^*} - \frac{S_y^2}{N}, \tag{3.4}$$

(Mukhopadhyay, 1998)

where the expected cost is

$$E(C) = C^* = c_1 n^{'} + n^{'} \sum c_h g_h W_h , \tag{3.5}$$

and $g_{h(opt)} = \left[\dfrac{c_1}{\left(S_y^2 - \sum W_h S_{yh}^2\right) c_h}\right]^{1/2} S_{yh}$; and hence

$$E\left(n_h^{''}\right) = n^{'} W_h g_{h(opt)} . \tag{3.6}$$

Pradhan (2000) considered a ratio estimator under two phase stratified sampling scheme given by

$$\overline{\overline{y}}_{Rst} = \sum_{h=1}^{L} w_h \frac{\overline{y}_h^{''}}{\overline{x}_h^{''}} \overline{x}_h^{'} \tag{3.7}$$

Here, a preliminary simple random sample without replacement (SRSWOR) sample s_1 of fixed size $n^{'}$ is selected and then classified into different strata with $n_h^{'}$ units falling in the h^{th} stratum $s_{1h}\left(h=1,2,\ldots,L\right)$ with $\sum_{h=1}^{L} n_h^{'} = n'$. In the second phase a SRSWOR sample s_{2h} of size $n_h^{''}$ is drawn from s_{1h} of size $n_h^{'}$ independently of each h to observe the main variable y. $\bar{\bar{y}}_{Rst}$ is approximately unbiased estimator of $\bar{Y}$ and the large sample variance of $\bar{\bar{y}}_{Rst}$ is given by

$$V\left(\bar{\bar{y}}_{Rst}\right)=\left(\frac{1}{n^{'}}-\frac{1}{N}\right)S_y^2+\sum_{h=1}^{L}\left(\frac{1}{g_h}-1\right)\frac{W_h S_{rh}^2}{n^{'}} \tag{3.8}$$

Using the cost function

$$C = c_1 n^{'} + \sum_{h=1}^{L} c_h\, n_h^{''} \tag{3.9}$$

the optimum value of the variance of $\bar{\bar{y}}_{Rst}$ is obtained as

$$V\left(\bar{\bar{y}}_{Rst}\right)_{opt}=\frac{\left[\sqrt{\left(S_y^2-\sum_{h=1}^{L}W_h S_{rh}^2\right)c_1}+\sum_{h=1}^{L}W_h S_{rh}\sqrt{c_h}\right]^2}{C^*}-\frac{S_y^2}{N} \tag{3.10}$$

where the expected cost is

$$E\left(C\right)=C^* = c_1 n^{'} + n^{'}\sum_{h=1}^{L} c_h\, g_h^* W_h, \tag{3.11}$$

$$g_{h(opt)}^{*} = \left[\frac{c_1}{\left(S_y^2-\sum W_h S_{rh}^2\right)c_h}\right]^{1/2} S_{rh}\text{, and hence}$$

$$E\left(n_h^{''}\right)=n^{'}W_h g_{h(opt)}^{*}. \tag{3.12}$$

4. THE SAMPLING DESIGN

Suppose a large sample s_1 of fixed size $n^{'}$ is drawn from a population of size N and is stratified on the basis of the stratifying variable Z. Let $n_h^{'}$ denote the number of units in $s_1\left(n^{'}\right)$ falling into h^{th} stratum $\left(h = 1, 2, \ldots, L, \sum_{h=1}^{L} n_h^{'} = n^{'}\right)$. A sub-sample $s_{2h}\left(n_h^{''}\right)$ is drawn from $s_{1h}\left(n_h^{'}\right)$ independently for each h and an auxiliary variable x is observed whose frequency distribution is unknown. Further, in the third phase a sub-sample $s_{3h}\left(n_h^{'''}\right)$ is drawn from $s_{2h}\left(n_h^{''}\right)$ independently for each h and the character of interest y, that is, the study variable, is observed. In the present study, simple random without replacement samples are selected in all the three phases.

5. PROPOSED RATIO ESTIMATOR IN THREE PHASE SAMPLING WITH STRATIFICATION

Define an estimator of population mean $\bar{Y}$ of y under three phase sampling set up given by

$$\overline{\overline{\overline{y}}}_{\mathrm{R}st} = \sum_{h=1}^{L} w_h \frac{\bar{y}_h^{'''}}{\bar{x}_h^{'''}} \bar{x}_h^{''} \tag{5.1}$$

THEOREM 5.1. $\overline{\overline{\overline{y}}}_{\mathrm{R}st}$ given by (5.1) is approximately an unbiased estimator of $\bar{Y}$.

PROOF. $E\left(\overline{\overline{\overline{y}}}_{\mathrm{R}st}\right) = E_1 E_2 E_3\left[\sum_{h=1}^{L} w_h \frac{\bar{y}_h^{'''}}{\bar{x}_h^{'''}} \bar{x}_h^{''}\right]$, where E_1, E_2, and E_3 denote expectation operators taken with respect to the first phase, second phase and third phase samples respectively.

$$= E_1 E_2\left[\sum_{h=1}^{L} w_h \bar{x}_h^{''} E_3\left(\frac{\bar{y}_h^{'''}}{\bar{x}_h^{'''}}\right)\right]$$

$\cong E_1 E_2\left[\sum_{h=1}^{L} w_h \bar{y}_h^{''}\right]$, since $E_3\left(\frac{\bar{y}_h^{'''}}{\bar{x}_h^{'''}}\right) \cong \frac{\bar{y}_h^{''}}{\bar{x}_h^{''}}$ neglecting bias of $o\left(1/n^{''}\right)$ for large value of $n_h^{'''}$

$$= E_1 \sum_{h=1}^{L} w_h \bar{y}_h^{'} = \bar{Y}$$

Thus, $\overline{\overline{\overline{y}}}_{Rst}$ is approximately unbiased estimator of $\overline{Y}$.

THEOREM 5.2. *If a first sample is a random sub-sample of fixed size $n^{'}$, the second sample is a random stratified sample from the first with fixed g_h $\left(o < g_h \le 1\right)$ and the third sample is a random sample from the second with fixed t_h $\left(o < t_h \le 1\right)$, then*

$$V\left(\overline{\overline{\overline{y}}}_{Rst}\right) \cong \left(\frac{1}{n^{'}} - \frac{1}{N}\right)S_y^2 + \sum_{h=1}^{L}\left(\frac{1}{g_h} - 1\right)\frac{W_h S_{yh}^2}{n^{'}} + \sum_{h=1}^{L}\frac{1}{g_h}\left(\frac{1}{t_h} - 1\right)\frac{W_h S_{rh}^2}{n^{'}} \tag{5.2}$$

PROOF. Now, $V\left(\overline{\overline{\overline{y}}}_{Rst}\right) = E_1E_2V_3\left(\overline{\overline{\overline{y}}}_{Rst}\right) + E_1V_2E_3\left(\overline{\overline{\overline{y}}}_{Rst}\right) + V_1E_2E_3\left(\overline{\overline{\overline{y}}}_{Rst}\right)$

(i)
$$\begin{aligned} E_1E_2V_3\left(\overline{\overline{\overline{y}}}_{Rst}\right) &= E_1E_2V_3\left[\sum w_h \frac{\overline{y}_h^{'''}}{\overline{x}_h^{'''}}\overline{x}_h^{''}\right] \\ &\cong E_1E_2\left[\sum w_h^2\left\{\frac{1}{n_h^{'''}} - \frac{1}{n_h^{''}}\right\}s_{rh}^{''2}\right] \\ &= E_1E_2\left[\sum w_h^2\left\{\frac{1}{n^{'}w_h}\cdot\frac{1}{g_h}\left(\frac{1}{t_h} - 1\right)\right\}s_{rh}^{''2}\right] \\ &\cong E_1\left[\sum_{h=1}^{L}\frac{\frac{1}{g_h}\left(\frac{1}{t_h} - 1\right)w_h s_{rh}^{'2}}{n^{'}}\right] \\ &= \sum_{h=1}^{L}\frac{\frac{1}{g_h}\left(\frac{1}{t_h} - 1\right)W_h S_{rh}^2}{n^{'}} \end{aligned} \tag{5.3}$$

since, $E\left(w_h s_{rh}^{'2}\right) = E\left\{w_h s_{rh}^{'2} \middle| n_h^{'}\right\} = E\left[w_h S_{rh}^2\right] = W_h S_{rh}^2$

(ii)
$$E_1V_2E_3\left(\overline{\overline{\overline{y}}}_{\mathrm{R}st}\right)=E_1V_2E_3\left[\sum_{h=1}^{L}w_h\frac{\overline{y}_h'''}{\overline{x}_h'''}\overline{x}_h''\right]$$
$$\cong E_1V_2\left[\sum_{h=1}^{L}w_h\overline{y}_h''\right]=E_1\left[\sum_{h=1}^{L}w_h^2\left(\frac{1}{n_h''}-\frac{1}{n_h'}\right)s_{yh}'^2\right]$$
$$\cong E_1\left[\sum_{h=1}^{L}\frac{\left(\frac{1}{g_h}-1\right)}{n'}w_h s_{yh}'^2\right]=\sum_{h=1}^{L}\frac{\left(\frac{1}{g_h}-1\right)}{n'}W_hS_{yh}^2 \tag{5.4}$$

(iii)
$$V_1E_2E_3\left(\overline{\overline{\overline{y}}}_{\mathrm{R}st}\right)=V_1E_2E_3\left[\sum w_h\frac{\overline{y}_h'''}{\overline{x}_h'''}\overline{x}_h''\right]$$
$$\cong V_1E_2\sum w_h\overline{y}_h''=V_1\left[\sum_{h=1}^{L}w_h\overline{y}_h'\right]=\left(\frac{1}{n'}-\frac{1}{N}\right)S_y^2 \tag{5.5}$$

Hence,

$$V\left(\overline{\overline{\overline{y}}}_{\mathrm{R}st}\right)\cong\left(\frac{1}{n'}-\frac{1}{N}\right)S_y^2+\sum_{h=1}^{L}\left(\frac{1}{g_h}-1\right)\frac{W_hS_{yh}^2}{n'}+\sum_{h=1}^{L}\frac{1}{g_h}\left(\frac{1}{t_h}-1\right)\frac{W_hS_{rh}^2}{n'} \tag{5.6}$$

THEOREM 5.3. An unbiased estimator of $V\left(\overline{\overline{\overline{y}}}_{\mathrm{R}st}\right)$ *is given by*

$$\hat{V}\left(\overline{\overline{\overline{y}}}_{\mathrm{R}st}\right)=\frac{1}{Nn'}\left[\left(\frac{N-1}{n'-1}\right)\sum_{h=1}^{L}n_h'\left(\frac{1}{g_h}-1\right)\right]+\frac{N-1}{n'-1}\sum_{h=1}^{L}\frac{1}{g_h}\left(\frac{1}{t_h}-1\right)n_h's_{rh}^2$$
$$+\frac{N-n'}{n'-1}\left\{\sum_{h=1}^{L}\frac{1}{g_ht_h}\sum_{j=1}^{n_h'''}y_{hj}^2-n'\overline{\overline{\overline{y}}}_{\mathrm{R}st}^2\right\} \tag{5.7}$$

PROOF. Using $Est.\{.\}$ as an estimator operator, we have the estimator of $V\left(\overline{\overline{\overline{y}}}_{\mathrm{R}st}\right)$ given by

$$Est.\,V\left(\overline{\overline{\overline{y}}}_{Rst}\right)=\hat{V}\left(\overline{\overline{\overline{y}}}_{Rst}\right)=Est.\left(\frac{1}{n'}-\frac{1}{N}\right)S_y^2+Est.\sum_{h=1}^{L}\left(\frac{1}{g_h}-1\right)\frac{W_hS_{yh}^2}{n'}$$
$$+Est.\sum_{h=1}^{L}\frac{1}{g_h}\left(\frac{1}{t_h}-1\right)\frac{W_hS_{rh}^2}{n'} \tag{5.8}$$

Now, $(N-1)S_y^2=\sum_{h=1}^{L}\sum_{j=1}^{N_h}y_{hj}^2-N\overline{Y}^2$

$$\therefore Est.\,(N-1)S_y^2=Est.\sum_{h=1}^{L}\sum_{j=1}^{N_h}y_{hj}^2-N\ Est.\,\overline{Y}^2$$
$$=Est.\sum_{h=1}^{L}\sum_{j=1}^{N_h}y_{hj}^2-N\left[\overline{\overline{\overline{y}}}_{Rst}^2-\hat{V}\left(\overline{\overline{\overline{y}}}_{Rst}\right)\right] \tag{5.9}$$

$$E\left[\sum_{h=1}^{L}\frac{w_h}{n_h'''}\sum_{j=1}^{n_h'''}y_{hj}^2\right]=\sum\frac{w_h}{N_h}\sum_{j=1}^{N_h}y_{hj}^2=\sum_{h=1}^{L}\frac{W_h}{N_h}\sum_{j=1}^{N_h}y_{hj}^2$$
$$=\sum_{h=1}^{L}\frac{N_h}{N}\cdot\frac{1}{N_h}\sum_{j=1}^{N_h}y_{hj}^2=\frac{1}{N}\sum_{h=1}^{L}\sum_{j=1}^{N_h}y_{hj}^2 \tag{5.10}$$

Hence

$$Est.\,(N-1)S_y^2==N\left[\sum_{h=1}^{L}\frac{w_h}{n_h'''}\sum_{j=1}^{n_h'''}y_{hj}^2-\left\{\overline{\overline{\overline{y}}}_{Rst}^2-\hat{V}\left(\overline{\overline{\overline{y}}}_{Rst}\right)\right\}\right] \tag{5.11}$$

$$Est.\sum_{h=1}^{L}\left(\frac{1}{g_h}-1\right)\frac{W_hS_{yh}^2}{n'}=\sum_{h=1}^{L}\left(\frac{1}{g_h}-1\right)\frac{w_hs_{yh}^2}{n'} \tag{5.12}$$

$$Est.\sum_{h=1}^{L}\frac{1}{g_h}\left(\frac{1}{t_h}-1\right)\frac{W_hS_{rh}^2}{n'}=\sum_{h=1}^{L}\frac{1}{g_h}\left(\frac{1}{t_h}-1\right)\frac{w_hs_{rh}^2}{n'} \tag{5.13}$$

The result (5.7) is found from (5.11), (5.12) and (5.13).

6. OPTIMUM ALLOCATION

Consider the cost function

$$C = c^{'}n^{'} + \sum_{h=1}^{L} c_h^{''} n_h^{''} + \sum_{h=1}^{L} c_h^{'''} n_h^{'''} \tag{6.1}$$

where,

$c^{'}$ = cost of observing a unit in the first phase

$c_h^{''}$ = cost of observing x -variate on a unit in the h^{th} stratum in the second phase

$c_h^{'''}$ = cost of observing y -variate on a unit in the h^{th} stratum in the third phase.

Here $n_h^{''}$ and $n_h^{'''}$ are random variables, hence we have

$$E\left(n_h^{''}\right) = E\left(g_h n_h^{'}\right) = g_h E\left(n_h^{'}\right) = n^{'} W_h g_h \tag{6.2}$$

$$E\left(n_h^{'''}\right) = E\left(t_h n_h^{''}\right) = t_h E\left(n_h^{''}\right) = n^{'} W_h g_h t_h \tag{6.3}$$

So the expected cost is given by

$$E(C) = C^{*}\,(say) = c^{'}n^{'} + n^{'}\sum_{h} c_h^{''} g_h W_h + n^{'}\sum_{h} c_h^{'''} g_h t_h W_h \tag{6.4}$$

It is required to find $n^{'}$, g_h and t_h so as to minimize $V\left(\bar{\bar{\bar{y}}}_{Rst}\right)$ for a given expected cost.

This is same as to minimize the product

$$\begin{aligned}
C^{*}\left[V + \frac{S_y^2}{N}\right] &= \left[c^{'} + \sum_{h=1}^{L} c_h^{''} g_h W_h + \sum_{h=1}^{L} c_h^{'''} g_h t_h W_h\right] \\
&\quad \times \left[S_y^2 + \sum_{h=1}^{L}\left(\frac{1}{g_h} - 1\right) W_h S_{yh}^2 + \sum_{h=1}^{L} \frac{1}{g_h}\left(\frac{1}{t_h} - 1\right) W_h S_{rh}^2\right] \\
&= \left[c^{'} + \sum_{h=1}^{L} c_h^{''} g_h W_h + \sum_{h=1}^{L} c_h^{'''} g_h t_h W_h\right] \\
&\quad \times \left[\left[S_y^2 - \sum_{h=1}^{L} W_h S_{yh}^2\right] + \sum_{h=1}^{L} \frac{W_h\left(S_{yh}^2 - S_{rh}^2\right)}{g_h} + \sum_{h=1}^{L} \frac{W_h S_{rh}^2}{g_h t_h}\right]
\end{aligned}$$

with respect to g_h and $g_h t_h$.

By applying Cauchy – Schwartz inequality, the minimum value of $C^*\left[V+\frac{S_y^2}{N}\right]$ occurs if and only if

$$\frac{c'}{S_y^2-\sum_{h=1}^{L}W_hS_{yh}^2}=\frac{c_h''g_hW_h}{\frac{W_h\left(S_{yh}^2-S_{rh}^2\right)}{g_h}}=\frac{c_h'''g_ht_hW_h}{\frac{W_hS_{rh}^2}{g_ht_h}}$$

This gives the optimum value of g_h and t_h as

$$g_{h(opt)}=\left[\frac{c'\left(S_{yh}^2-S_{rh}^2\right)}{c_h''\left(S_y^2-\sum_{h=1}^{L}W_hS_{yh}^2\right)}\right]^{1/2} \quad \text{and} \quad t_{h(opt)}=\frac{S_{rh}\sqrt{c_h''}}{\sqrt{\left(S_{yh}^2-S_{rh}^2\right)c_h'''}} \tag{6.5}$$

Substituting the value of n', $g_{h(opt)}$ and $t_{h(opt)}$ in the variance expression, the optimum variance is obtained as

$$V^*=V\left(\bar{\bar{\bar{y}}}_{Rst}\right)_{opt}$$

$$=\left[\sqrt{c'\left\{S_y^2-\sum_{h=1}^{L}W_hS_{yh}^2\right\}}+\sum_{h=1}^{L}W_h\sqrt{\left(S_{yh}^2-S_{rh}^2\right)c_h''}+\sum_{h=1}^{L}W_hS_{rh}\sqrt{c_h'''}\right]^2C^{*-1} \tag{6.6}$$

$$-\frac{S_y^2}{N}$$

7. NUMERICAL ILLUSTRATION

Example - 7.1

Consider a hypothetical population given in Table 1, where population $N=250$ units are distributed in $L=2$ strata.

Let the expected cost of the experiment be $C^*=60$. If the cost per unit of observation is $c=1.5$, then for SRSWOR a sample size $n=40$ is permissible.

Now, $S_y^2 = \sum_{h=1}^{2} W_h S_{yh}^2 + \sum_{h=1}^{2} W_h \left(\bar{Y}_h - \bar{Y}\right)^2 = 52.4176$.

Hence, $V\left(\bar{y}_{ran}\right) = \left(\frac{1}{n} - \frac{1}{N}\right) S_y^2 = 1.1008$.

TABLE 1

Strata	N_h	S_{yh}^2	S_{xh}^2	$\bar{Y}_h$	$\bar{X}_h$	R_h	ρ_h
1	160	25	36	30	40	0.75	0.90
2	90	9	16	18	25	0.72	0.85

As $S_{rh}^2 = S_{yh}^2 + R_h^2 S_{xh}^2 - 2R_h \rho_h S_{yh} S_{xh}$, we have $S_{r1}^2 = 4.75$ and $S_{r2}^2 = 2.6064$.

For $c_h = c = 1.5$ the optimum variance of $\bar{\bar{y}}_{st}$ in two phase stratified sampling with cost function $C = c_1 n' + \sum_{h=1}^{2} c_h n_h''$ is given by (1.3) i.e.,

$$V\left(\bar{\bar{y}}_{st}\right)_{opt} = \frac{\left(5.76\sqrt{c_1} + 5.2419\right)^2}{60} - 0.2097.$$

Now $V\left(\bar{\bar{y}}_{st}\right)_{opt} < V\left(\bar{y}_{ran}\right)$, if $c_1 < 0.3962$.

For $c_h = c = 1.5$, the optimum variance of $\bar{\bar{y}}_{st}$ in two phase stratified sampling with ratio method of estimation with the cost function $c = c_1 n' + \sum_{h=1}^{2} c_h n_h''$ is given by (1.7) i.e.

$$V\left(\bar{\bar{y}}_{Rst}\right)_{opt} = \frac{\left(6.9598\sqrt{c_1} + 2.4202\right)^2}{60} - 0.2097.$$

Now, $V\left(\bar{\bar{\bar{y}}}_{Rst}\right)_{opt} < V\left(\bar{\bar{y}}_{st}\right)_{opt}$ if $c_1 < 0.8581$.

Setting $c_1 = 0.25$ and $c_h = 1.5$, we find $V\left(\bar{\bar{y}}_{st}\right)_{opt} = 0.8897$ and $V\left(\bar{\bar{y}}_{Rst}\right)_{opt} = 0.3705$.

For $c_h'' = c_1 = 0.25,\ c_h''' = c_h = 1.5$, the optimum variance of $\bar{\bar{\bar{y}}}_{Rst}$ in three phase stratified sampling with ratio method of estimation is given by (5.6) i.e.,

$$V\left(\bar{\bar{\bar{y}}}_{Rst}\right)_{opt} = \frac{\left(5.76\sqrt{c'} + 4.31535\right)^2}{60} - 0.2097$$

$V\left(\bar{\bar{\bar{y}}}_{Rst}\right)_{opt} < V\left(\bar{\bar{y}}_{st}\right)_{opt}$ if $c' < 0.4367$

$V\left(\bar{\bar{\bar{y}}}_{st}\right)_{opt} < V\left(\bar{\bar{y}}_{Rst}\right)_{opt}$ if $c' < 0.0757$.

Finally, setting $c^{'} = 0.05, c_h^{''} = 0.25,\ c_h^{'''} = 1.5$, we find $V\left(\overline{\overline{\overline{y}}}_{\mathrm{R}st}\right)_{opt} = 0.3136$.

Hence, the relative precision of the various methods can be summarized as follows:

TABLE 2

	Sampling Method	Method of Estimation	Relative Precision
1.	Simple Random	Mean per unit	100
2.	Stratified random	Two phase	123.7271
3.	Stratified random	Two phase ratio	297.1120
4.	Stratified random	Three phase ratio	351.0204

The optimum value of the sampling fractions for two phase stratified random sampling are given by $g_{1(opt)} = 0.35438$ and $g_{2(opt)} = 0.21263$.

From the expected cost we find $n^{'} = 85,\ E\left(n_1^{''}\right) = 19,\ E\left(n_2^{''}\right) = 12$.

The optimum value of the sampling fractions for two phase stratified random sampling with ratio method of estimation are given by $g_{1(opt)} = 0.12784$ and $g_{2(opt)} = 0.09470$.

From the expected cost we find $n^{'} = 142,\ E\left(n_1^{''}\right) = 12,\ E\left(n_2^{''}\right) = 5$.

The optimum value of the sampling fractions for three phase stratified random sampling with ratio method of estimation are given by $g_{1(opt)} = 0.34939$, $g_{2(opt)} = 0.19632,\ t_{1(opt)} = 0.19772,\ t_{2(opt)} = 0.26066$.

From the expected cost we find

$n^{'} = 276,\ E\left(n_1^{''}\right) = 62,\ E\left(n_2^{''}\right) = 20,\ E\left(n_1^{'''}\right) = 12,\ E\left(n_2^{'''}\right) = 5$.

Note: $n^{'}$ is determined by the expected cost, $c^{'}, c_h^{''}, c_h^{'''}, W_h$, $g_{1(opt)}, g_{2(opt)}, t_{1(opt)}, t_{2(opt)}$ and may be more than N in a given situation. As $n^{'} = 276$ is more than the finite population size $N = 250$, the first phase sample size is restricted to 250, resulting in a two phase sampling design.

Example - 7.2

The following data come from a particular census in a given year, of all farms in Jefferson County, Iowa. In this example, y_{hj} represents area in acres under corn and x_{hj} as total area in acres in the farm. The population is divided into two strata, the first stratum containing farms of area upto 160 acres and the second stratum containing farms of more than 160 acres.

Let the expected cost of the experiment be $C^{*} = 50$.

If the cost per unit of observation is $c = 0.5$, then for SRSWOR a sample size $n = 100$ is permissible.

TABLE 3

Strata	Size (Acres)	N_h	S_{yh}^2	S_{xh}^2	$\bar{Y}_h$	$\bar{X}_h$	R_h	ρ_h
1	0 – 160	1580	312	2055	19.404	82.56	0.2350	0.61694
2	more than 160	430	922	7357	51.626	244.85	0.2109	0.32945

Hence, $V(\bar{y}_{ran}) = 5.86474$.

As $S_{rh}^2 = S_{yh}^2 + R_h^2 S_{xh}^2 - 2R_h \rho_h S_{yh} S_{xh}$, we have $S_{r1}^2 = 193.3078,\ S_{r2}^2 = 887.3112$.

For $c_h = c = 0.05$, the optimum variance of $\bar{\bar{y}}_{st}$ in two phase stratified sampling with the cost function $C = c_1 n' + \sum_{h=1}^{2} c_h n_h''$ is given by (1.3) i.e.,

$$V(\bar{\bar{y}}_{st})_{opt} = \frac{\left(13.2167\sqrt{c_1} + 20.3806\sqrt{c_h}\right)^2}{50} - 0.30705.$$

Now, $V(\bar{\bar{y}}_{st})_{opt} < V(\bar{y}_{ran})$ if $c_1 < 0.057$.

For $c_h = c = 0.05$, the optimum variance of $\bar{\bar{y}}_{st}$ in two phase stratified sampling with ratio method of estimation with the cost function $C = c_1 n' + \sum_{h=1}^{2} c_h n_h''$ is given by (1.7) i.e.

$$V(\bar{\bar{y}}_{Rst})_{opt} = \frac{\left(16.5953\sqrt{c_1} + 17.30164\sqrt{c_h}\right)^2}{50} - 0.30705.$$

Now, $V(\bar{\bar{y}}_{Rst})_{opt} < V(\bar{y}_{ran})$ if $c_1 < 0.103$.

Setting $c_1 = 0.015,\ c_h = 0.5,$ we find $V(\bar{\bar{y}}_{st})_{opt} = 4.8322$ and $V(\bar{\bar{y}}_{Rst})_{opt} = 3.7637.$

For $c_h'' = c_1 = 0.015$ and $c_h''' = c_h = 0.5$, the optimum variance of $\bar{\bar{\bar{y}}}_{st}$ in three phase stratified sampling with ratio method of estimation is given by (5.6) i.e.,

$$V\left(\bar{\bar{\bar{y}}}_{Rst}\right)_{opt} = \frac{\left(13.2167\sqrt{c'} + 13.4373\right)^2}{50} - 0.30705.$$

Now, $V\left(\bar{\bar{\bar{y}}}_{Rst}\right)_{opt} < V(\bar{\bar{y}}_{st})_{opt}$ if $c' < 0.03848$ and $V\left(\bar{\bar{\bar{y}}}_{Rst}\right)_{opt} < V(\bar{\bar{y}}_{Rst})_{opt}$ if $c' < 0.0039$.

Finally, setting $c' = 0.003,\ c_h'' = 0.015,\ c_h''' = 0.5$, we find $V\left(\bar{\bar{\bar{y}}}_{Rst}\right)_{opt} = 3.7037$.

Hence, the relative precision of the various methods can be summarized as follows:

The optimum value of the sampling fractions for two phase stratified random sampling are given by $g_{1(opt)} = 0.23148$, $g_{2(opt)} = 0.39793$

TABLE 4

	Sampling Method	Method of Estimation	Relative Precision
1.	Simple Random	Mean per unit	100
2.	Stratified random	Two phase	121.3679
3.	Stratified random	Two phase ratio	155.8238
4.	Stratified random	Three phase ratio	158.3481

From the expected cost we find $n' = 337$, $E\left(n_1''\right) = 61$, $E\left(n_2''\right) = 29$.

The optimum value of the sampling fractions for two phase stratified random sampling with ratio method of estimation are given by $g_{1(opt)} = 0.14511$, $g_{2(opt)} = 0.31090$.

From the expected cost we find $n' = 475$, $E\left(n_1''\right) = 54$, $E\left(n_2''\right) = 32$.

The optimum value of the sampling fractions for three phase stratified random sampling with ratio method of estimation are given by $g_{1(opt)} = 0.36864$, $g_{2(opt)} = 0.19929$ and $t_{1(opt)} = 0.22104$, $t_{2(opt)} = 0.87600$.

From the expected cost we find $n' = 852$, $E\left(n_1''\right) = 246$, $E\left(n_2''\right) = 36$, $E\left(n_1'''\right) = 55$, $E\left(n_2'''\right) = 32$.

8. CONCLUSIONS

The three phase stratified sampling with ratio estimator is of use when the strata weights are unknown and the mean of the auxiliary variables in different strata are unknown. The comparison of optimum variance of stratified two phase sampling with known strata weights and with ratio method of estimation with the optimum variance of stratified three phase sampling with unknown strata weights and unknown population means of auxiliary variable in different strata are theoretically intractable. However, the given numerical illustrations show that there might arise situations with specific cost function, when the three phase stratified sampling with ratio method of estimation is more efficient than the stratified two phase sampling and the stratified two phase sampling with ratio method of estimation with known strata weights. The proposed method is based on large sample approximations and extensive simulation studies with small sample sizes would be useful for better appreciation of results obtained in this paper.

ACKNOWLEDGEMENTS

Author is thankful to Prof. A.K.P.C. Swain for his kind suggestions. The author also wishes to express sincere gratitude to the referee for his valuable suggestions in improving the manuscript.

REFERENCES

W.G. COCHRAN (1977), *Sampling Technique* (Third Edition), John Wiley and Sons, New York.

A.L. FINKNER (1950), *Methods of Sampling for estimating commercial peach production in North Carolina*, North Carolina Agr. Exp. Stat. Tech. Bull. 91

P. MUKHOPADHYAY (1998), *Theory and Methods of Survey Sampling*, Prentice Hall of India, New Delhi.

J.N. PASCUAL (1961), *Unbiased Ratio Estimators in Stratified Sampling*, Jour. Amer. Stat. Assoc., 56, 80 – 87.

B.K. PRADHAN (2000), *Some problems of Estimation in Multi-phase Sampling*, Thesis submitted for the degree of Ph.D degree in Statistics of the Utkal University, Bhubaneswar.

B.K. PRADHAN, A.K.P.C. SWAIN (2000), *Two Phase Stratified Sampling with Ratio and Regression Methods of Estimations*, Journal of Science and Technology, Vol. XII, Section B, 52 – 57.

J.N.K. RAO (1973), *On Double Sampling for stratification and analytical surveys*, Biametrica, 60,125 – 133.

D.S. ROBSON (1952), *Multiple Sampling of Attributes*, Jour. Amer. Stat. Assoc., 47, 203 – 215.

D.S. ROBSON, A.J. KING (1953), *Double Sampling and the Curtis impact Surveys*, Cornell Univ. Expt. Stat. Mem., 231.

A.K.P.C. SWAIN (1973), *Some Contributions to the Theory of Sampling*, Thesis submitted for the degree of Ph.D degree in Statistics of the Utkal University, Bhubaneswar.

SUMMARY

Three phase stratified sampling with ratio method of estimation

In this paper a three phase sampling is proposed to estimate the population mean of character by a three phase stratified ratio estimator and some numerical results are presented to illustrate efficiency of the proposed procedure against possible alternative ones and from the expected cost optimum sample sizes on first, second and third phases are also calculated.