Thailand Statistician
January 2026; 24(1): 116-123
http://statassoc.or.th
Contributed paper

Parameter Estimation in Case of Incomplete Frames using Ratio Estimators

Shilpa Yadav* **and Pankaj Nagar**
Department of Statistics, University of Rajasthan, Jaipur, Rajasthan, India
*Corresponding author; e-mail: shilpa.y2108@gmail.com

Received: 27 August 2023
Revised: 10 July 2024
Accepted: 4 August 2025

Abstract

Existence of a well defined and perfect sampling frame is the fundamental requirement of any sample survey. But there is enough evidence to support the fact that a perfect sampling frame that captures all the individual units of the population is rarely available, especially for a dynamic population where a constant movement of the units of the population is observed. In such cases, the sample collected can not be considered as a good representative of the population and the problem of incomplete frame arises. The results so drawn can immensely change the survey results and influence the legitimacy of the research. This study deals with the incomplete frame problem using ratio method of estimation, where the information on the auxiliary or ancillary variable is collected in the first phase and a second phase sample is then drawn to obtain estimates of the population mean of the characteristic under study and its mean square error up to first order approximation. Two different estimators, viz., combined ratio estimator and separate ratio estimators have been used and their efficiencies are compared. Further, the results are illustrated numerically with the help of Monte-Carlo simulations.

Keywords: Incomplete sampling frame, double sampling, ratio method of estimation, predecessor-successor method, ancillary variable

1. Introduction

Data collection is often considered as the indispensable phase of any sample survey and it requires a well defined sampling frame that is perfect in all sense. A perfect sampling frame is the one that consists of all the units of the population such that each unit gets a fair chance of inclusion in the sample while a survey is been conducted. But quite often this chance or probability of inclusion of the sampling units is compromised due to numerous factors, incompleteness in sampling frame being the most common one. Incompleteness in a sampling frame leads to the problem of undercoverage bias (Bethlehem and Biffignandi 2012) and it occurs when some units or a certain segment of the target population are entirely excluded in the sampling frame. Such units can be referred to as 'unincluded units'. When this type of frame problem arises, the fundamental assumption of any sampling scheme is violated and the units unincluded in the frame receive a zero probability of inclusion in the sample. This makes the conventional Neyman's probability sampling paradigm inapplicable.

An example of incomplete frame or undercoverage as raised by Bethlehem and Biffignandi (2012) is a survey where the sample is taken from a population register. In such surveys, illegal immigrants may be considered as a part of the population, but since they will have no official record in the population register, they will never be a part of the sample.

Yates (1948) in his first edition, discussed in detail, about the various imperfections that a sampling frame can suffer from. He mentioned incompleteness of sampling frame as the most common weakness of a sampling frame. Hartley (1962) made use of multiple frames for dealing with the problem of incomplete sampling frames. In dual or multiple frame technique, two or more frames (each considered as incomplete individually) are used simultaneously for sampling and composite estimators are then formed on basis of the two samples. Later, many authors as Burmeister (1972), Saxena, Narain and Srivastava (1984) and others took upon the approach of dual and multiple frames in order to simplify the problem of incomplete sampling frames. Later on, Hansen, Hurwitz, and Jabine (1963), gave the method of predecessor-successor for dealing with the problem of incomplete sampling frames, when a geographical ordering of the sampling units is possible and thus each unit can be traced by following a certain path of travel on the basis of its predecessor or successor. In this method, every unit sampled is used to check whether the unit next to it occurs in the frame, if yes, then they surveyor should move to the next unit sampled, otherwise the next unit excluded from the frame is included in the sample and the process is continued till he encounters with the unit already existing in the sampling frame. This method is of great importance while dealing with the problem of missing units or incomplete sampling frame. Many authors (Singh (1983), Singh (1989), Agarwal and Gupta (2008), Gupta et al. (2021), Joshi et al. (2021), Yadav et al. (2022) and others) used this technique for development of estimators of population mean and total, based on included as well as the unincluded units under different sampling schemes.

While conducting a sample survey, a surveyor is confronted with all kinds of phenomena that may lead to inefficient estimates of the population characteristic under study. Many a times, two common causes of sampling errors are confused with one another, viz., incomplete sampling frame and non-response. Where the prior is a problem of sampling frame, i.e.,it occurs when there are certain units which are missing from the frame, the latter refers to the survey error and occurs when the information on the unit is missing. The two errors are completely different and occur at two different phases of a sample survey. But a common error they cause, is the inefficiency of the estimates produced in the presence of these problems. There have been continuous efforts to improve the efficiency of estimators so that more reliable estimates can be obtained for the population parameters.

In order to improve the efficiency of the population parameter estimates, there are a number of sampling techniques that necessitates additional information in the form of one or more ancillary or auxiliary variable, say X, to increase the efficiency of the estimator for estimating the population mean or total of the variable under study, say Y. Some of these techiques being ratio estimator developed by Cochran (1940), product estimator developed by Murthy (1964) and regression method of estimation developed by Hansen, Hurwitz and Madow (1953). Ratio and product methods of estimation are preferred when the regression line of Y on X passes through the origin and there is strong positive and negative correlation, respectively, between the variable under study and ancillary variable. Regression method provides a relationship between the two variables, when the line of regression of Y on X does not pass through origin. But, these sampling schemes, require the knowledge about the population mean of the ancillary variable in order to estimate the population mean and its variance of the variable under study. Though, in many practical situations, where the existing sampling frame is found to be incomplete, the available information on the ancillary variable may not be suitable, as it can not be considered adhering to all the sampling units in the population.

A rescue to such problem may be double sampling, which can effectively be used for tracing the unincluded units in the first phase and using the information so collected in the second phase for the purpose of estimation. In this scheme, a primary sample of relatively larger size is selected from the known population and the data is collected with respect to the ancillary variable. While doing so, the units that occur preceeding or succeeding the sampled units are also marked and the data for ancillary variable is thus noted. Thus, at the end of this exercise, there are two samples, one that was originally selected from the existing frame and another that was traced using the predecessor-successor method. An estimator of population mean of ancillary variable is obtained from this first phase samples. Then, the two first phase samples are further used for second phase sampling and data on the variable under study is recorded. Finally, estimate of population mean and its variance can be obtained using ratio, product or regression estimators.

2. Statement of the Problem

In most of the countries, population censuses are conducted once every 10 years. Meanwhile, before the next census, lot of changes take place in the system which are not registered and included in the existing sampling frame. If a specific survey is conducted for a specific reason after, say 4 years, the existing sampling frame becomes inadequate and incomplete for the current situation. Therefore, the available sampling frame, if used for estimating any characteristic, provides vague results and will not be suitable for any further use. To obtain better estimates of the characteristic under study, one needs to inculcate the new information added in the forms of new units existing in population.

Let us consider, a survey for estimating the total sales of goods and services in a specific locality in India, where Department of Labour of every state keeps record of shops and establishments. For the survey, a sampling frame of all the registered shops is obtained by the information gathered from the Department of Labour, where the Shop and Establishment Act regulates the process of registration for businesses. But as soon as surveyor visits the local market, he observes that there are some unregistered shops being operated under residential holdings. These shops, which are unregistered, are contributing to the sales purchase data but are not actually counted as they are not a part of the available sampling frame. The existing sampling frame would undercover the actual population and provide vague results. Thus, it becomes of paramount importance to include information from those shops to generate more precise and accurate estimates of the total sales.

In countries like India, there is a common problem of incompleteness in sampling frames, as the construction of a perfect sampling frame is still a big challenge. Lack of technology, under-covered population and inadequate resources are major reasons behind that. Despite being such a deep rooted sampling frame error, incompleteness in sampling frames is still neglected and under-studied. This study is an attempt to generate a solution to this problem alongwith providing improved results.

Let us consider a finite population $U = \{U_1, U_2, ..., U_{N_1}\}$ that consists of N_1 units listed in the sampling frame which is to be used for sampling and further estimation. Let Y be the variable under study and X be the ancillary variable that is strongly positively correlated with the variable under study. From the available sampling frame, a sample of size n_1' is selected using simple random sampling without replacement (SRSWOR) scheme and is then surveyed for information on the ancillary variable. While doing so, it is observed that there are certain units which are present in the field and should necessarily be a part of the study, but they are somehow excluded in the frame, let them be called “unincluded units”. So, the existing frame becomes incomplete and inadequate to be used for the study, as a portion of the population, that of unincluded units, will always remain unattended and neglected, giving rise to inefficient and vague results. In this situation, ordinary ratio method of estimation with simple random sampling without replacement scheme, fails, as the data available on the population total or average of ancillary variable, is incomplete.

Here, for field surveys, it can be assumed that there is a geographical ordering of the sampling units, such that all the unincluded units exist randomly between the included units as -

$$U_1 \text{——} M_1;\ U_2 \text{——} M_2; ...;\ U_{N_1} \text{——} M_{N_1}.$$

where U_i denotes the i^{th} unit of the population and M_i denotes the number of unincluded units lying between the i^{th} and $(i+1)^{th}$ included unit. Let there be a total of N_2 such unincluded units that can be traced with the help of their succeeding or preceeding included unit using Predecessor-Successor method given by Hansen, Hurwitz, and Jabine (1963). Here, by definition, we have $\sum_{i=1}^{N_1} M_i = N_2; 0 \leq M_i \leq N_2$.

Now, let a sample of n_2' unincluded units is obtained while enumerating n_1' included units and data relating to X is recorded, so that a total on n units are observed in the first phase. Now there are two samples, one that of included units(n_1') and another of unincluded units(n_2'). Here, it can easily be shown that an estimate of total unincluded units ($\hat{N}_2$) is given as $\hat{N}_2 = \frac{n'_2}{n'_1} N_1$. The two samples so obtained, can be considered as two strata from which, further sampling is done in the second phase for studying the variable Y. A sample of size n is selected from the first phase samples using SRSWOR, where n_1 units can be considered to have been come from n'_1 first phase units (fsu) and another sample of size n_2 to have been selected from n'_2 first phase units. These two samples are now surveyed for Y and data is collected.

Here, the simple double sampling ratio estimator would only be based on the existing sampling frame (only N_1 units), and thus the estimate would be

$$\hat{\bar{Y}}_{Rd1} = \frac{\bar{y}_1}{\bar{x}_1} \cdot \bar{x}'_1. \tag{1}$$

The mean square error, here is obtained as

$$MSE(\hat{\bar{Y}}_{Rd1}) = \left(\frac{1}{n_1} - \frac{1}{N_1}\right) S^2_{y_1} + \left(\frac{1}{n_1} - \frac{1}{n'_1}\right) \left(R_1^2 S^2_{x_1} - 2\rho_1 R_1 S_{x_1} S_{y_1}\right). \tag{2}$$

This method would totally exclude all the sampling units that were not a part of existing sampling frame, thus providing incomplete and unreliable results. Here, including the unincluded, the double sampling ratio estimator can be obtained in two different approaches, viz. combined and separate ratio estimators.

2.1. Combined ratio estimator

Here, a double sampling combined ratio estimator is defined as

$$\hat{\bar{Y}}_{Rdc} = \frac{\bar{y}}{\bar{x}} \cdot \bar{x}' = \hat{R} \cdot \bar{x}'$$

where

$$\begin{aligned}
&\hat{R} = r = \frac{\bar{y}}{\bar{x}};\ R = \frac{\bar{Y}}{\bar{X}} \\
&\bar{x}' = w_1{}'\bar{x}'_1 + w_2{}'\bar{x}'_2;\ \bar{y} = w_1\bar{y}_1 + w_2\bar{y}_2;\ \bar{x} = w_1\bar{x}_1 + w_2\bar{x}_2 \\
&w_1{}' = \frac{n_1{}'}{n_1{}' + n_2{}'};\ w_2{}' = \frac{n_2{}'}{n_1{}' + n_2{}'};\ w_1 = \frac{n_1}{n_1 + n_2};\ w_2 = \frac{n_2}{n_1 + n_2} \\
&E(w_1) = E(w'_1) = W_1;\ E(w_2) = E(w'_2) = W_2.
\end{aligned} \tag{3}$$

W_1 and W_2 being the population proportions of the included and unincluded units.
$\bar{x}'$ is the weighted sample mean of the ancillary variable in the first phase and $\bar{y}$ and $\bar{x}$ are the weighted sample means of the variable under study and ancillary variable in the second phase respectively. $\bar{y}$ being an unbiased estimate of population mean of variable under study $\bar{Y}$ and $\bar{x}'$ and $\bar{x}$ being unbiased estimates of population mean of the ancillary variable $\bar{X}$. Here, it is practically very difficult to obtain the exact expressions for $E(\bar{y}/\bar{x})$ and $E(\bar{y}^2/\bar{x}^2)$, so we use approximation of first order to find the bias and the mean square error (MSE) of $\hat{\bar{Y}}_{Rdc}$.

2.1.1 Properties of combined ratio estimator

Bias

A large sample approximation of second order gives

$$Bias(\hat{\bar{Y}}_{Rdc}) = E(\hat{\bar{Y}}_{Rdc}) - \bar{Y} \approxeq \sum_{i=1}^{2} E(w_i^2) R S_{xi} \left[\left(\frac{1}{n_i} - \frac{1}{n'_i}\right)\left(\frac{S_{xi}}{\bar{X}} - \frac{\rho_i S_{yi}}{\bar{Y}}\right)\right] \tag{4}$$

where $w_i; i = 1, 2$ are the weights associated such that

$$E(w_i^2) = V(w_i) + W_i^2 = \left(\frac{N-n'}{N-1}\frac{W_i(1-W_i)}{n'}\right) + W_i^2;\ N = N_1 + N_2, n' = n_1' + n_2'.$$

$\bar{X}$ and $\bar{Y}$ are the weighted population means of the two samples. R is the population ratio estimator with $\rho_i; i = 1,2$ being the coefficient of correlation between the ancillary variable and variable under study in two strata respectively.

S_{xi} and S_{yi} are the population sample standard deviations of the ancillary variable and variable under study respectively, for the i^{th} stratum.

Mean Square Error

The mean square error of the double sampling combined ratio estimator up to second order of approximation is obtained as

$$\begin{aligned} MSE(\hat{\hat{Y}}_{Rdc}) =& E\left[\hat{\hat{Y}}_{Rdc} - \bar{Y}\right]^2 \\ \approxeq& \sum_{i=1}^{2}\left[\left(V(w_i)\bar{Y}_i^2 + \left(V(w_i) + W_i^2\right)\right)\right. \\ & \left.\left\{\left(\frac{1}{n_i} - \frac{1}{N_i}\right)S_{yi}^2 + \left(\frac{1}{n_i} - \frac{1}{n_i'}\right)\left(R^2 S_{xi}^2 - 2\rho_i R S_{xi} S_{yi}\right)\right\}\right] \end{aligned}$$

where $V(w_i) = \dfrac{N-n'}{N-1}\dfrac{W_i(1-W_i)}{n'}$,

since $w_1 + w_2 = 1$, $V(w_1) = \dfrac{N-n'}{N-1}\dfrac{W_1 W_2}{n'} = V(w_2)$; S_{xi}^2 and S_{yi}^2 having their usual meaning. An estimate of MSE of $\hat{\hat{Y}}_{Rdc}$ is given as

$$\begin{aligned} \widehat{MSE}(\hat{\hat{Y}}_{Rdc}) \approxeq& \sum_{i=1}^{2}\left[\left(\hat{V}(w_i)\bar{y}_i^2 + \left(\hat{V}(w_i) + w_i^2\right)\right)\right. \\ & \left.\left\{\left(\frac{1}{n_i} - \frac{1}{N_i}\right)s_{yi}^2 + \left(\frac{1}{n_i} - \frac{1}{n_i'}\right)\left(r^2 s_{xi}^2 - 2\rho_i r s_{xi} s_{yi}\right)\right\}\right] \\ \hat{V}(w_i) =& \frac{N-n'}{N-1}\frac{w_i(1-w_i)}{n'} \end{aligned}$$

where $r = {}^{\bar{y}}/_{\bar{x}}$ is the sample ratio estimator, which is an unbiased estimator of $R = {}^{\bar{Y}}/_{\bar{X}}$.

2.2. Separate ratio estimator

Considering the entire population, as a constitution of two strata of sizes N_1 (known) and N_2 (unknown) containing all included and unincluded units, respectively. The estimates of population mean of ancillary variable for two strata is obtained in the first phase as

$$\bar{x}_1' = \hat{\hat{X}}_1 = \frac{1}{n_1'}\sum_{r=1}^{n_1'}\bar{x}_r; \quad \bar{x}_2' = \hat{\hat{X}}_2 = \frac{1}{n_2'}\sum_{s=1}^{n_2'}\bar{x}_s \tag{5}$$

n_1' and n_2' being the number of included and unincluded (traced using Predecessor-Successor method) units in the first phase.

The two strata so formed are used for further sampling. A double sampling separate ratio estimator is hereby defined as

$$\hat{\hat{Y}}_{Rds} = w_1\bar{y}_{Rd1} + w_2\bar{y}_{Rd2} \tag{6}$$

where $w_i; i = 1,2$ are same as in (3)

$$\bar{y}_{Rd1} = \frac{\bar{y}_1}{\bar{x}_1}\cdot\bar{x}_1'; \quad \bar{y}_{Rd2} = \frac{\bar{y}_2}{\bar{x}_2}\cdot\bar{x}_2'.$$

2.2.1 Properties of separate ratio estimator

Bias

A large sample approximation of second order gives, the bias of the separate ratio estimate as

$$\begin{aligned} Bias(\hat{\hat{Y}}_{Rds}) =& E\left(w_1\bar{y}_{Rd1} + w_2\bar{y}_{Rd2}\right) \\ =& \sum_{i=1}^{2} W_i\bar{Y}_i \left(\frac{1}{n_i} - \frac{1}{n_i'}\right)\left(C_{xi}^2 - \rho_i C_{xi}C_{yi}\right), \end{aligned} \quad (7)$$

where $w_i; i = 1, 2$ are the weights associated and ρ_i is again the coefficient of correlation between the ancillary and variable under study in two strata. Here, $C_{xi}; i = 1, 2$ and $C_{yi}; i = 1, 2$ are the coefficient of variation for ancillary and variable under study respectively. $\bar{Y}_i$ is the population mean of the variable under study.

Mean Square Error

The mean square error of the double sampling separate ratio estimator up to second order approximation is obtained as

$$\begin{aligned} MSE(\hat{\hat{Y}}_{Rds}) =& \sum_{i=1}^{2}\left[\left(V(w_i)\bar{Y}_i^2 + \left(V(w_i) + W_i^2\right)\right)\right. \\ & \left.\left\{\left(\frac{1}{n_i} - \frac{1}{N_i}\right)S_{yi}^2 + \left(\frac{1}{n_i} - \frac{1}{n_i'}\right)\left(R_i^2 S_{xi}^2 - 2\rho_i R_i S_{xi} S_{yi}\right)\right\}\right] \end{aligned} \quad (8)$$

where $V(w_i) = \dfrac{N - n'}{N - 1}\dfrac{W_i(1 - W_i)}{n'}$,

since $w_1 + w_2 = 1$, $V(w_1) = \dfrac{N - n'}{N - 1}\dfrac{W_1 W_2}{n'} = V(w_2)$; S_{xi}^2 and S_{yi}^2 having their usual meaning.

An estimate of MSE, in this case, is obtained as

$$\begin{aligned} \widehat{MSE}(\hat{\hat{Y}}_{Rds}) =& \sum_{i=1}^{2}\left[\left(\hat{V}(w_i)\bar{y}_i^2 + \left(\hat{V}(w_i) + w_i^2\right)\right)\right. \\ & \left.\left\{\left(\frac{1}{n_i} - \frac{1}{N_i}\right)s_{yi}^2 + \left(\frac{1}{n_i} - \frac{1}{n_i'}\right)\left(r_i^2 s_{xi}^2 - 2\rho_i r_i s_{xi} s_{yi}\right)\right\}\right] \\ \hat{V}(w_i) =& \frac{N - n'}{N - 1}\frac{w_i(1 - w_i)}{n'}. \end{aligned}$$

3. Numerical Illustration

This study is an attempt to improve the estimates of the population mean and its MSE while dealing with the problem of incomplete sampling frames by using additional information associated with the variable under study in the form of ancillary variable. To demonstrate the usage and efficacy of the estimator that have been developed in this study, the authors have used the "MU284" dataset in R, given by Sarndal, Swenson, and Wretman (1992). The data consists of 11 variables corresponding to 284 municipalities, out of which the authors have tried to estimate the mean revenue obtained from the various municipalities of Sweden in 1985 given as RMT85 (Revenues from 1985 municipal taxation, in millions of kronor) as the study variable and the population of the corresponding municipality (defined under variable P85) as the ancillary variable, both variables having high positive correlation ($\rho = 0.98$). As the literature does not provide data related to incomplete frames in the form, authors desired, the MU284 population is divided into two sub-populations, by marking the unincluded units (N_2, in aggregate) randomly, depending upon the population proportion(PP_{inc}, ranging from 0.5 to 0.9), of the included units(N_1, in aggregate). A first phase sample of the included units is then selected using SRSWOR of size n_1' from the N_1 units , based on sample proportion for first phase, SP_{inc1} that also ranges from 0.6 to 0.8. On the basis of the sample produced, n_2' units are traced using Predecessor-Successor method. The data on ancillary variable is then recorded. Further, two samples (size n_1 and n_2) are selected in the second phase from the two first phase samples of included and unincluded units, respectively, based on the sample proportion for second phase SP_{inc2} (ranging from 0.1 to 0.3). The data on the variable under study, viz. RMT85, is then collected and used for estimation. The population mean of the variable under study, its bias along with its MSE using simple double sampling ratio estimator, double sampling combined ratio estimator and double sampling separate ratio estimator is calculated. Apparantely, there is no algebraic comparison of the so obtained mean square errors feasible. To solve this problem, simulated results are used to calculate percentage relative efficiency (PRE) of the estimators. The percentage relative efficiency is calculated as $PRE = \frac{MSE_1}{MSE_2} \times 100$, MSE_1 and MSE_2 being the respective mean square errors of the two estimators. Since, enumeration of all possible combinations of first phase and second phase samples is not possible, the authors have used Monte-Carlo simulation to support the claims of the theory developed in this study.

The authors performed 10000 simulations with varying levels of PP_{inc}, SP_{inc1} and SP_{inc2}. The estimates of number of unincluded units were obtained thereafter along with mean, bias, mean square error and relative efficiencies of the three different estimators (simple double sampling ratio estimator (say, Rd), double sampling combined ratio estimator (Rdc, say) and double sampling separate ratio estimator (Rds, say)). For illustration, the results obtained for included population proportion 0.9, 0.7 and 0.5 are tabulated as support in 1. Also, very accurate estimates of the unincluded population size were obtained. Results showed a constant increase in the percentage relative efficiency as the included population proportion decreased from 0.9 to 0.5.

Table 1 Percentage Relative Efficiencies at various levels of PP_{inc}, SP_{inc1} and SP_{inc2}

SP_{inc1}	SP_{inc2}	$PP_{inc} = 0.9$			$PP_{inc} = 0.7$			$PP_{inc} = 0.5$		
		PRE1	PRE2	PRE3	PRE1	PRE2	PRE3	PRE1	PRE2	PRE3
0.8	0.1	**114.72**	**120.02**	**104.62**	**165.76**	**194.34**	**117.24**	107.29	**160.63**	**149.72**
	0.2	**118.32**	**121.51**	**102.69**	**177.23**	**193.88**	**109.39**	**132.23**	**164.69**	**124.55**
	0.3	**119.44**	**121.63**	**101.83**	**181.74**	**192.73**	**106.05**	**141.23**	**162.02**	**114.72**
0.7	0.1	**118.37**	120.26	**101.59**	166.25	191.06	114.93	121.36	171.74	141.52
	0.2	**118.63**	121.25	102.21	176.61	190.77	108.01	142.19	172.73	121.48
	0.3	118.94	120.78	**101.54**	179.71	189.06	105.20	146.53	164.80	112.47
0.6	0.1	117.65	119.11	101.24	**165.77**	187.86	113.32	137.29	**181.78**	132.41
	0.2	119.00	120.67	101.40	174.57	186.86	107.04	145.57	173.28	119.04
	0.3	124.00	**126.65**	102.14	181.62	**189.61**	104.40	**152.00**	**168.72**	**111.00**

$PRE1$ refers to the PRE of Rdc over Rd, $PRE2$ is the PRE of Rds over Rd and $PRE3$ is the PRE of Rds over Rdc

4. Conclusions

Results show that the double sampling combined and separate ratio estimator perform better than the usual double sampling estimator for all $0.5 \leq PP_{inc} \leq 0.9$. The consistency in the efficiency is observed for different proportions of sampling in the first and second phase as well. Also, results support the fact that double sampling separate ratio estimator generates better estimates of population characteristic with lesser standard error as compared to the other two. For $0.5 \leq PP_{inc} \leq 0.9$, the proposed methods provide refined estimates of population characteristics without constructing a complete sampling frame again from scratch. For $PP_{inc} = 0.5$ or less, construction of a fresh sampling frame for estimation purpose is suggested as that indicates a lot number of units being missed from the existing frame. The concept can further be utilised for estimation using simple random sampling with replacement scheme and other complex sampling techniques too.

Acknowledgements

We are deeply indebted to Prof. P. C. Gupta for his constant support and guidance.

References

Agarwal B and Gupta PC. Estimation from incomplete sampling frames in case of simple random sampling. Model Assist Stat Appl. 2008; 3(2): 497-516. doi 10.3233/MAS-2008-3204

Bethlehem J, Biffignandi S. The Problem of Undercoverage. Handbook of Web Surveys. New York: John Wiley & Sons; 2012.

Burmeister LF. Estimators for samples selected from multiple overlapping frames. PhD [dissertation]. Iowa State University; 1972.

Cochran WG. The Estimation of the Yields of cereal experiments by sampling for the ratio of grain in total produce. J Agric Sci. 1940; 30(2); 262-275.

Gupta PC, Joshi L, Joshi V, Nagar P. Weighted product estimator for incomplete sampling frame. J Rajasthan Acad Phy. 2019; 18(3-4): 233-240.

Gupta PC, Joshi V, Nagar P, Singh AK. Linear regression estimator in case of incomplete sampling frame. J Rajasthan Acad Phys Sci. 2021; 20(1-2): 49-56.

Hansen MH, Hurwitz WN, Jabine TB. The use of imperfect lists for probability sampling at United-States Bureau of Census. B Int Statist Inst. 1963; 40(1): 497-516.

Hartley HO. Multiple frame surveys. Proceedings of the Social Statistics Section, American Statistical Association. 1962; 203-206.

Hansen MH, Hurwitz WN, Madow WG. Use of incomplete frame in large scale sample survey. New York: John Wiley & Sons; 1953.

Joshi V, Nagar P, Singh AK, Gupta PC. Use of ratio method of estimation in incomplete frames. Int J Agric Stat Sci. 2021; 17: 267-272.

Kish L. Survey Sampling. New York: John Wiley & Sons; 1965.

Lohr S, Rao JNK. Estimation in multiple-frame surveys. J Am Stat Assoc. 2006; 101(475): 1019-1030.

Murthy MN. Product method of estimation. Sankhya Ser A. 1964; 26; 69-74.

Sarndal CE, Swenson B, Wretman J. Model assisted survey sampling. New York: Springer-Verlag; 1992.

Saxena BC, Narain P, Srivastava AK. Multiple frame surveys in two stage Sampling. Sankhya Ser B. 1984; 46; 75-82.

Seal KC. Use of out-dated frames in large scale sample surveys. Calcutta Stat Assoc. 1962; 11(3): 68-84. doi 10.1177/0008068319620302

Singh D, Chaudhary FS. Theory and Analysis of Sample Survey Designs. New York: Wiley; 1986.

Singh NK, Kumar R, Sehgal VK. Use of incomplete frame in large scale sample survey. Gujarat Stat Rev. 2001; 28; 3-10.

Singh R. On the use of incomplete frames in sample surveys. Biometrical J. 1983; 25(6): 545-549. doi:10.1002/bimj.19830250605

Singh R. Method of estimation for sampling from incomplete frames. Aust J Stat. 1989; 31(2): 269-276. 10.1111/j.1467-842x.1989.tb00396.x

Yadav S, Bundel R, Singh AK, Nagar P. Parameter estimation in case of incomplete sampling frames: a simulation approach. J Rajasthan Acad Phys. 2022; 21; 255-266.

Yates F. Sampling methods for census and surveys. J R Stat Soc Ser A-G. 1949; 112(4): 483-484.