Statistics and Applications {ISSN 2454-7395 (online)}
Volume 21, No. 2, 2023 (New Series), pp 107–120
http://www.ssca.org.in/journal

Dealing with the Imperfect Frame Arising Due to Rare Outdated Units from Finite Population

Neelam Kumar Singh
Brahmanand Post Graduate College, Rath (Hamirpur), UP, India

Received: 01 October 2021; Revised: 26 April 2022; Accepted: 15 November 2022

Abstract

Frame is quite often incomplete and imperfect by the time actual survey starts. Imperfection of frame also arises due to some of the rare units being out-of–scope, out-dated and missing from the sampling frame which may be of considerable measure in size and weight and may lead to deviation between sampled and target population. The estimators from such imperfect frame will not give unbiased results. Unbiased estimator is being devised for target population total and variance in the present study with appropriate sampling design considering the finite population. It is considered that, in the case of finite population, the phenomenon of some of the rare units missing, out of scope or out-dated units follow some probability distribution function (p.d.f) and unbiased estimate of population parameter is devised and developed.

Key words: Imperfect frame; Target population; Inverse sampling.

1. Introduction

In some situations, there are some rare units in the sampled population which are out-dated from the target population at the time of actual survey. The rare units of the old frame for which sampled population correspond, may be out-dated, may be out-of-scope or may be missing from the frame at the time of actual survey for which the information and observations are desired. Such rare units missing from the target population will lead to the imperfection of the frame. Thus there is deviation between sampled population and target population due to imperfection in the frame arising due to some rare sampling units being out-dated from the frame required for the study of statistical results of the desired target population. Therefore, in such cases some rare units of the sampled population do not belong to the target population because frame prepared at some time may contain some rare units which may not exist in the target population at the time of actual survey and enumeration. This happens because during passage of time, the frame prepared at some time will be out-dated by the time the actual enumeration starts.

For example in a frame of list of Agriculture labourers in a district, some rare number

Corresponding Author: Neelam Kumar Singh
Email: neelamkumar099@gmail.com

of labourers may migrate to other districts in search of work at the time, the survey actually starts. Therefore, frame of Agricultural labourers become imperfect at the time of enquiry.

The list of irrigated land holdings in a Tehesil prepared at some time corresponding to the sampled population, may contain some rare number of fields which may not in-fact be irrigated discovered at the time of actual enquiry of crop survey due to failure of some canals or tube wells. In a list of cultivated area under some crop designed for the purpose of survey from source at hand, may contain some cultivated area under some other crop so that some of the rare number of cultivated area may not belong to the crop-area desired for the study. The frame becomes imperfect due to some rare sampling units not belonging to the target population which are unknown at the time of sample selection but is discovered because investigator visits particular cultivated land selected in the sample.

There are also some rare phenomena in the nature when some rare units change rapidly in their geographic ordering and location. For example, in case of shifting cultivation adopted in North-East of India, some of the rare area listed under forest land may be found to be under shifting cultivation due to unprecedental customs and traditions of tribes in that region because of dependence of tribes on nature and nurture. They are discovered only at the time of investigation. Similarly, frame of fields under Jhooming cultivation obtained from some source may contain some rare field which may be discovered as no more belonging under same pattern of agriculture at the time of actual enumeration as is evident in number of Anthropological studies.

The frame, available for some nomadic tribal families will become rapidly imperfect because of some rare number of nomadic tribal families migrating from one place to other, thus making their demographic studies difficult and complex at the time of actual survey. In a frame of mango trees for estimation of total amount of mango fruits, some rare number of trees may not be found bearing fruits discovered at the time of enquiry, In a list of fields for tomato, potato, wheat, gram, Arahar etc, some rare field may contain damaged crops due to adverse weather conditions and estimation of total production of these crops would be difficult on the basis of available old frame as frame become imperfect for desired target population under study. For estimating total amount of Tendu leaves in a forest, list of Tendu tree may contain rare trees which may not bear leaves during passage of time making the frame imperfect. For estimating total amount of water in a district, the list of water tanks available from some official sources, may not correspond to the target population as some rare tanks may be dry and barren by the time of survey.

Thus more often, frame may very soon become out–dated as some rare units of the frame may go quantitative and qualitative change in the sampled population.

Estimates on the basis of sample selected from such imperfect frame would not give unbiased results. These rare out-dated units from the frame will also contribute to the bias of target population results. The correct, complete, and up-to-date frame is rather impossible in practice because some rare units of the sampled population rapidly cease to exist or are observed as non- existent in the target population at the time of actual survey.

Seal (1962) discussed the use of out-dated frames in large scale surveys and considered the changes in the population as a continuous stochastic process. Hartley (1962) proposed the use of two or more frames to overcome the problem of incomplete frames. Hansen *et* *al.* (1963) discussed various procedures for the use of incomplete frame and proposed the predecessor-successor method to obtain information on missing units in the frame. Szamsitat and Schaffer (1963) discussed about consequences of imperfect frame in sampling. Singh (1983) gave a mathematical formulation of predecessor-successor method for estimating total number of units missing from the frame. Singh (1989) proposed suitable method of estimation when sampling is done from imperfect frame and a geographic ordering of units can be established. Singh *et al.* (1997) discussed imperfection in the frame of finite population and proposed estimators for domain of study considering probability distribution of the out-dated units in the incomplete frame. Singh *et al.* (2001) discussed the imperfection of frame arising due to omission of some of the units from the frame and also frame containing some units which no more belongs to target population and proposed appropriate estimation procedure for the population, its variance when sampling is done from two frames.

Agarwal and Gupta (2008) developed a method of estimation of population total, mean and variance from incomplete frame in case of SRSWOR and SRSWR.

Singh (2020) discussed the frame error as error due to imperfection of the frame in detail because of deviation of the target population from sample population. Singh (2020) discussed that frames are often imperfect in any sample survey which arises due to some of the rare sampling units being out-dated at the time of actual survey. He further devised unbiased estimator for the imperfection of frame arising due to rare out-of-scope units considering population size to be large. Suitable estimators for the proportion of out-dated units from the population and target population total for a character with their variance was developed considering probability distribution function (p.d.f.) of rare out- dated units in large population.

Agarwal and Singh (2021) considered every finite of population as a constituent part of some super population in which complete frame is quite often not available under consideration. They providers estimators under more realistic situation as compared to finite population concepts.

Singh (2021) discussed that the existence of the frame is pre-requisite for any sample survey or census of a large population. Frames are quite often imperfect due to dynamic nature of sampling units. Frames become incomplete by the time actual survey and enumeration starts which affect the statistical results desired for the target population. He reported and considered imperfection in the frame of large population arising due to the qualitative change of units from one class to other. He considered incomplete frame assuming the nature of units following dynamic change from class one to other which follows a probability distribution function. Suitable estimator for proportion of units belonging to a particular domain and unbiased estimate of target population for a class was proposed along with its estimate of variance. The estimates are evolved so as to eliminate error caused due to the deviation of sampled population from target population.

Singh (2022) gave appraisal of problem of incomplete frame in different situations. The cause and type of imperfection of frame was discussed covering various aspects and review of work done by different scholars was explained along with measures and suggestions to deal with imperfection of frame.

More troublesome are those cases when missing or out-dated units although rare in number are exceedingly large in their measure of size and such rare units were discovered because units were selected. For example, in list of business establishments, some large establishments although rare in number, may no longer be active at the time of enumeration.

These large sized rare units missing from the frame would lead to the imperfection of the frame and would contribute much to the bias of sampling results for target population parameters.

Therefore, this study deals with imperfect frame arising due to missing or out- dated rare units from the target population of finite size. The objective in the present study is to design a sampling procedure to devise an estimator which is unbiased for target population parameters in such cases of imperfection in the frame.

2. Method of estimation

Consider a finite sampled population of size N as listed in the available frame for selecting the sample at some time. However, during passage of time, population structure has undergone some change in the sense that some rare units of the available original frame *i.e.* of sampled population, have ceased to exist in the target population. Let N_1 denotes the number of units in the frame of size N, actually belonging to the target population. N_2 denotes number of rare sampling units which have gone out of the target population. So that, $N = (N_1 + N_2)$. The N_2 units, though rare in the number may attribute for high measure of their size. Assume that the rare units which have ceased to exist in the target population are not identifiable and hence these rare units although being out-of scope and out-dated units, cannot be deleted from the frame. Therefore, actual frame of target population will be of size N_1 which is unknown at the time of actual survey. N_2 units are rare in number but these are also not identifiable which are attributing for their high measure of sizes with considerable importance and significant for population parameters in their observations of characteristics under study.

Situation of incomplete frame arises when units under go qualitative change and becomes out-dated units for the desired targeted population, during passage of time. The units respond but their measures/ values are out-dated for target population. In case of non-response, frame is complete but some of the units do not respond.

The information and observations attributed to the N_2 rare units of the sampled population are, although, of significant importance but will not correspond the characteristics of the target population. This phenomenon occurs when there are changes in the target population when population is subject to continuous change and if there is rather long interval between the dates to which sampled population relates and date or time for which the information is to be collected.

One procedure to select the sample may be to select a random number from 1 to N keeping old numbering as such. The unit corresponding to this number is selected, provided it is not of N_2 rare units which have become non-existent in the target population. If non-existent rare unit is selected, draw is rejected and the procedure is repeated. This gives equal chance of selection to $N - N_2 = N_1$ units of the target population.

However, this procedure assumes that the information is available about the rare out-dated units from the old frame at the time of sample selection. But, most often, we do not know about the out-dated units from the old frame at the time of sample selection unless the actual enumeration starts. It is only when enumerator visits a particular rare out-dated unit that he finds that the units no more exist in the target population.

Therefore, we propose an alternative method of sample selection and sampling plan for estimation procedure.

3. Notations

Let p denotes the proportion of rare sampling units from the available original frame of sampled population which are out-dated, out-of-scope or missing units from the target population leading to the imperfection of the frame. Hence, $p = N_2/N$. Evidently Np units will be rare units which are missing from the target population. We have $N = N_1 + Np$ and $N_1 = Nq, q = 1 - p$. Hence, Nq units actually belong to the target population.

Let Y denotes the character under study and Y_i denote the value of Y for ith unit of the sampled population

$$U = (U_1, U_2, U_3, ..., U_N).$$

Let $\bar{Y}_1 = \frac{1}{N_1}\sum_{i=1}^{N_1} Y_i$: population mean of the target population

$$S_1^2 = \frac{1}{N_1}\sum_{i=1}^{N_1}(Y_i - \bar{Y}_{N_1})^2$$

$p = N_2/N$: proportion of rare sampling units in the sampled population which are non-existent in the target population.
$q = N_1/N$: proportion of sampling units in the available frame of sampled population actually belonging to the target population.

The population total for the target population can be easily written as

$$Y_{N_1} = N_1\bar{Y}_1.$$

In order to estimate target population total for Y, we have to estimate N_1 and $\bar{Y}_1$.

4. Proposed sampling procedure: Inverse simple random sampling without replacement (I.S.R.S.W.O.R.)

Since, the number of out-dated units is rare in number; the proportion p is very small. In such situation, method of Inverse-sampling can be used with advantage. In this method, the sampled size n is not fixed in advance. Instead sampling is continued until a predetermined number of rare units out-dated from the frame have been drawn. Let p denotes the proportion of rare units missing from the frame. Evidently Np units will be missing or out-dated units in the frame of sampled population and $N - Np = N_1$ units will exist in the target population so that $N_1 = Nq, q = 1 - p$. For estimating the population p, the sampling units are drawn one by one with equal probability of selection and without replacement. This procedure is called I.S.R.S.W.O.R. sampling is discontinued as soon as the number of units in the sample possessing the rare units missing from the frame is some predetermined number.

In some situations, the statistical investigation may demand a sample to include a required representation from the category of rare missing units. It may be required of a sample to include a specified number of rare units which are out-dated. In such situation direct sampling procedure based on fixed sample size may not be appropriate. Rather, inverse sampling procedure discussed by Haldane (1945), Finney (1949), Chapman (1952) and Chikkagudar (1969) among others are expected to be more appropriate. Suppose, the statistical enquiry requires that the sample should include n_2 units from the rare missing units from the frame.

We continue selecting units, one by one, with equal probability and without replacement from the sampled population $U = \{U_1, U_2, ..., U_N\}$ until there are exactly n_2 units (given) discovered at the enumeration stage represented from the missing units. The total number of units, n in the sample is obviously a random variable. Method of continuing sample, called I.S.R.S.W.O.R has one important advantage. This case of I.S.R.S.W.O.R can be used with advantage when proportion $p \leq 10\%$ (Haldane, 1945). In such situation p is small but not well known in advance. The value of n is large if p is small. Thus sample of size n will contain n_2 units which do not exist in the target population and $n - n_2 = n_1$ (say) units belong in the target population after enumerator visits each unit of the sample of size n. As such, there can be no observations obtained for such n_2 (given) non-existent units in the frame. The observations for $n - n_2$ units can be obtained for which observations are available. The n_2 units are non-existent or even if they exist, the enumerator can identify them as not belonging to the target population and hence cannot be observed.

Therefore, corresponding probability distribution $P(n)$ for random variable n is given by

$$p(n) = p\left\{\begin{array}{c}\text{In a sample of } (n-1) \text{ units drawn} \\ \text{from } N, n_2 - 1 \text{ units will be} \\ \text{discovered to be missing or out-dated} \\ \text{units in the target population}\end{array}\right\}.p\left\{\begin{array}{c}\text{the unit drawn at the} \\ n\text{th draw will be rare} \\ \text{missing or out-dated unit} \\ \text{from the target population}\end{array}\right\}$$
$$= \frac{\binom{N_2}{n_2-1}\binom{N_1}{n-n_2}}{\binom{N}{n-1}}.\frac{N_2-(n_2-1)}{N-(n-1)}$$

where $n = n_2, n_2+1, ..., (n_2 + N - N_2)$
or

$$p(n) = \frac{\binom{Np}{n_2-1}\cdot\binom{Nq}{n-n_2}}{\binom{N}{n-1}}.\frac{Np-(n_2-1)}{N-(n-1)}$$

where $n = n_2, n_2+1, ..., n_2 + Nq$. Here, we also have $\sum p(n) = 1$, where $n \geq n_2$.

5. Estimation of proportion and its variance

It can be shown that an unbiased estimate of p can be given by

$$Est\ p = \frac{n_2 - 1}{n-1} = \hat{p},\ \text{say}. \tag{1}$$

Thus,

$$
\begin{aligned}
E\left(\frac{n_2-1}{n-1}\right) &= \sum\left(\frac{n_2-1}{n-1}\right)\times\frac{\binom{Np}{n_2-1}\cdot\binom{Nq}{n-n_2}}{N\binom{N-1}{n-1}}\cdot\frac{Np-n_2+1}{N-n+1}\\
&= \sum\frac{(n_2-1)}{(n-1)}\frac{(n-1)}{(n_2-1)}Np\frac{\binom{Np-1}{n_2-2}\cdot\binom{Nq}{n-n_2}}{N\binom{N-1}{n-2}}\frac{Np-n_2+1}{N-n+1}\\
&= p\sum_{n\geq n_2}\frac{\binom{Np-1}{n_2-2}\cdot\binom{Nq}{n-n_2}}{\binom{N-1}{n-2}}\frac{Np-n_2+1}{N-n+1}\\
&= p
\end{aligned}
$$

We shall now determine the estimate of $V(\hat{p})$. We know that

$$V(\hat{p}) = E(\hat{p}^2)-(E(\hat{p}))^2 = E(\hat{p}^2)-p^2. \tag{2}$$

Therefore, an biased estimate of $V(\hat{p})$ is given by

$$Est\ V(\hat{p}) = \hat{p}^2 - Est\ p^2$$

Now to determine $Est\ p^2$ we have

$$E\frac{(n_2-1)(n_2-2)}{(n-1)(n-2)} = \frac{Np(Np-1)}{N(N-1)}\sum_{n\geq n}\frac{\binom{Np-2}{n_2-3}\binom{Nq}{n-n_2}}{\binom{N-2}{n-3}}\frac{Np-n_2+1}{N-n+1} \quad = \frac{N}{N-1}p^2-\frac{1}{N-1}p$$

Therefore, $Est\ \frac{N}{N-1}p^2 - \frac{1}{N-1}Est\ p = \frac{(n_2-1)(n_2-2)}{(n-1)(n-2)}$ or,

$$Est\ p^2 = \frac{N-1}{N}\frac{(n_2-1)(n_2-2)}{(n-1)(n-2)}+\frac{1}{N}\hat{p}. \tag{3}$$

Therefore, from (2) and (3), we have

$$
\begin{aligned}
Est\ V(\hat{p}) &= \hat{p}^2-\frac{N-1}{N}\frac{(n_2-1)(n_2-2)}{(n-1)(n-2)}-\frac{1}{N}\hat{p}\\
&= \hat{p}\left\{\hat{p}-\left(\frac{N-1}{N}\right)\left(\frac{n_2-2}{n-2}\right)-\frac{1}{N}\right\}
\end{aligned} \tag{4}
$$

or,

$$\hat{V}(\hat{p}) = \frac{(n_2-1)^2}{(n-1)^2}-\frac{(N-1)(n_2-1)(n_2-1)}{N(n-1)(n-2)}-\frac{(n_2-1)}{N(n-2)} \tag{5}$$

Similarly $V(\hat{p})$ can be obtained as

$$V(\hat{p}) = \frac{(n_2-1)p+np(1-p)}{N-1} \tag{6}$$

Therefore, we have an unbiased estimate of $\hat{Y}_{N_1}$ (target population total), as given by $Est\ \hat{Y}_{N_1} = Est\ (Nq\bar{Y}_1) = Est\ (Nq)Est(\bar{Y}_1) = \hat{Y}_{N_1}$ say.
Thus,

$$\begin{aligned} \hat{Y}_{N_1} &= (\hat{Nq})(\hat{\bar{Y}}_1) = N\hat{q}\bar{y}_1 \\ &= N(1-\hat{p})\bar{y}_1 \\ &= N\left(\frac{n-n_2}{n-1}\right)\bar{y}_1 \end{aligned} \tag{7}$$

$\hat{Y}_{N_1}$ is an unbiased estimate of Y_{N_1}, because
$E(\hat{Y}_{N_1}) = E(N\hat{q}\bar{y}_1) = Nq\bar{Y}_1 =$ Target population total.

6. Variance of $\hat{Y}_{N_1}$

The variance of estimate of the target population total is given by

$$\begin{aligned} V(\hat{Y}_{N_1}) &= V(N\hat{q}\bar{y}_1) \\ &= N^2V\{(1-\hat{p})\bar{y}_1\} \\ &= N^2[V(\bar{y}_1) + E\{V(\hat{p}\bar{y}_1)|\hat{p}\} + V(E(\hat{p}\bar{y}_1|\hat{p}))] \\ &= N^2[V(\bar{y}_1) + E\{\hat{p}^2V(\bar{y}_1)\} + V(\bar{Y}_1\hat{p})] \\ &= N^2[V(\bar{y}_1) + V(\bar{y}_1)E(\hat{p}^2) + \bar{Y}_1^2V(\hat{p})] \end{aligned} \tag{8}$$

as $E(\bar{y}_1) = \bar{Y}_1$ and also $E(\hat{p}^2) = V(\hat{p}) + p^2$. Here $\bar{y}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} y_i$ which is sample mean for y based on n_1 observations and $n_1 = n - n_2$ for which observation are available in the sample belonging to the target population. Again number of units belonging to the target population are $Nq = N_1$ whose variance is given by S_1^2. Sample mean square error for units in the target population can be given by

$$s_1^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(y_i - \bar{y}_1)^2.$$

Also, $\bar{Y}_1 =$ Mean of the target population so that

$$\bar{Y}_1 = \frac{1}{N_1}\sum_{i=1}^{N_1} Y_i.$$

We can also have

$$V(\bar{y}_1) = \frac{N_1-n_1}{N_1n_1}S_1^2.$$

Thus, again we have from (8)

$$\begin{aligned} V(\hat{Y}_{N_1}) &= N^2[V(\bar{y}_1) + V(\bar{y}_1)\{V(\hat{p}) + \hat{p}^2\} + \bar{Y}_1^2V(\hat{p})] \\ &= N^2[V(\bar{y}_1)\{1 + \hat{p}^2 + V(\hat{p})\} + \bar{Y}_1^2V(\hat{p})]. \end{aligned} \tag{9}$$

Putting values of $V(\bar{y}_1)$ and $V(\hat{p})$ we can also have

$$\begin{aligned} V(\hat{Y}_{N_1}) = N^2 \left[\frac{N_1 - n_1}{N_1 n_1} S_1^2 \left\{ 1 + p^2 + \frac{(n_2 - 1)p + Np(1-p)}{N-1} \right\} \right] \\ + N^2 \bar{Y}_1^2 \left\{ \frac{(n_2 - 1)p + Np(1-p)}{N-1} \right\}. \end{aligned} \tag{10}$$

If $\frac{N}{N-1} \cong 1$ then we have after simplification

$$V(\hat{Y}_{N_1}) = \frac{N(N_1 - n_1)}{N_1 n_1} S_1^2 \{N - 1 + p(N + n_2 p - 1)\} + Np\bar{Y}_1^2 \{N(1-p) + n_2 - 1\} \tag{11}$$

and as $N_1 = Nq = N(1-p)$, we can also have

$$V(\hat{Y}_{N_1}) = \frac{N(1-p) - n_1}{(1-p)n_1} S_1^2 \{N - 1 + p(N + n_2 - p - 1)\} + Np\bar{Y}_1^2 \{N(1-p) + n_2 - 1\}. \tag{12}$$

Thus $V(\hat{Y}_{N_1})$ is function of N, p, n_1, S_1^2, n_2, and $\bar{Y}_1^2$. Second term in the $V(\hat{Y}_{N_1})$ is independent of n but first term is not independent of n because $n_1 = n - n_2$. For fixed n_2, n_1 increases as n (random variable) increases. Therefore, for a given n_2, $V(\hat{Y}_{N_1})$ *i.e.* variance of the estimate of target population total decreases as sample size n increases. However, $V(\hat{Y}_{N_1})$ increases as n decreases. But $n \geq n_2$ so that

$$n = n_2, n_2 + 1, n_2 + 2, ..., n_2 + N(1-p).$$

For $n_2 = n$, *i.e.*, when $n_1 = 0$, it is not possible to determine $V(\hat{Y}_{N_1})$.

However, if $n = n_2 + 1$ then $n_1 = 1$ and $V(\hat{Y}_{N_1})$ is maximum and is given by

$$V(\hat{Y}_{N_1}) = \frac{N(1-p) - 1}{((1-p)} S_1^2 \{N - 1 + p(N + n - p - 2)\} + Np\bar{Y}_1^2 \{N(1-p) + n_2 - 2\}.$$

If n is the maximum value which is given as $n = n_2 + N(1-p)$ then $n_2 = n - N(1-P)$ and $n_1 = N(1-p)$. In this case variance of $V(\hat{Y}_{N_1})$ is given by

$$V(\hat{Y}_{N_1}) = N(n-1)p\bar{Y}_1^2.$$

Again $V(\hat{Y}_{N_1})$ also depends on the nature of p. The behavior of the frame depends on the proportion p. It can be seen that, in case of perfect frame $p = 0$. Therefore, when there is no imperfection in the frame $V(\hat{Y}_{N_1}) = \frac{(N-n_1)}{n_1} S_1^2 (N-1)$, which is approximately equal to the $V(\hat{Y}_{N_1})$. In case of perfect frame, when there are no rare missing units in the sampled population. We have

$$V(\hat{Y}_{N_1}) = V(\hat{Y}_N) = \frac{(N - n_1)}{n_1}(N-1)S^2$$

as $S_1^2 = S^2$ and $n_1 = n$. The approximation occurs as we have assumed $\frac{N}{(N-1)} \cong 1$ in (11) earlier. However, if $p = 1$ *i.e.* when there is total imperfection in the frame then $V(\hat{Y}_{N_1})$ cannot be determined.

We know that $0 \leq p \leq 1$. It can be seen that $V(\hat{Y}_{N_1})$ is maximum as $p \to 0$.

7. Estimation of $V(\hat{Y}_{N_1})$

We have from (9)

$$V(\hat{Y}_{N_1}) = N^2[V(\bar{y}_1)(1+p^2) + V(\bar{p})\{V(\bar{y}_1) + \bar{Y}_1^2\}].$$

Estimate of $V(\hat{Y}_{N_1})$ can be obtained by estimating each of the right hand side terms. Therefore,

$$Est\ V(\hat{Y}_{N_1}) = N^2[Est\ V(\bar{y}_1)(1+p^2) + Est\ V(\bar{p}) Est\ \{V(\bar{y}_1) + \bar{Y}_1^2\}].$$

As we know that

$$V(\bar{y}_1^2) = E(\bar{y}_1^2) - \bar{Y}_1^2$$

therefore,

$$E(\bar{y}_1^2) = V(\bar{y}_1^2) + \bar{Y}_1^2$$

and

$$Est\ V(\bar{y}_1^2) + Est\ (\bar{Y}_1^2) = \bar{y}_1^2.$$

Similarly,

$$Est\ (1+p^2) = 1 + Est\ p^2.$$

But we know that

$$V(\hat{p}^2) = E(\hat{p}^2) - p^2$$

and

$$p^2 = E(\hat{p}^2) - V(\hat{p}).$$

Hence,

$$Est\ p^2 = \hat{p}^2 - \hat{V}(\hat{p}). \tag{13}$$

Putting these values in (9) we can obtain

$$\begin{aligned}\hat{V}(\hat{Y}_{N_1}) &= N^2\left[Est\ V(\bar{y}_1)\{1+\hat{p}^2 - \hat{V}(\hat{p})\} + Est\ V(\hat{p})(\bar{Y}_1^2)\right] \\ &= N^2\left[\frac{N_1 - n_1}{N_1 n_1} s_1^2\{1+\hat{p}^2 - \hat{V}(\hat{p})\} + \hat{V}(\hat{p})\bar{Y}_1^2\right].\end{aligned} \tag{14}$$

Since n_1 is also selected with S.R.S.W.O.R. in I.S.R.S.W.O.R. with n_2 fixed and n random so that $n_1 = n - n_2$ and we have $s_1^2 = \frac{1}{(n_1-1)}\sum_{i=1}^{n_1}(y_i - \bar{y}_1^2)$ Putting value of $\hat{V}(\hat{p})$ from equation (5) in (14), we have

$$\hat{V}(\hat{Y}_{N_1}) = N^2\left[\frac{N_1 - n_1}{N_1 n_1} s_1^2\left\{1 + \frac{N-1}{N}\hat{p}\left(\frac{n_2-2}{n-2}\right) + \frac{1}{N}\hat{p}\right\}\right] + \bar{y}_1^2\left\{\hat{p}^2 - \frac{N-1}{N}\hat{p}\left(\frac{n_2-2}{n-2}\right) - \frac{1}{N}\hat{p}\right\} \tag{15}$$

or

$$\hat{V}(\hat{Y}_{N_1}) = N\left\{\frac{N_1 - n_1}{N_1 n_1} s_1^2\right\}\left\{N + (N-1)\hat{p}\frac{n_2-2}{n-2} + \hat{p}\right\} + N\bar{y}_1^2\left\{N\hat{p}^2 - (N-1)\hat{p}\frac{n_2-2}{n-2} - \hat{p}\right\}. \tag{16}$$

8. Estimation of $\frac{1}{N_1}$

As N_1 is unknown, it has to be substituted with its estimate, therefore, we can have

$$\frac{N_1 - n_1}{N_1 n_1} = \left(\frac{1}{n_1} - \frac{1}{N_1}\right).$$

Thus,

$$Est\ \left(\frac{N_1 - n_1}{N_1 n_1}\right) = Est\ \left(\frac{1}{n_1} - \frac{1}{N_1}\right) = \frac{1}{n_1} - Est\ \frac{1}{N_1}.$$

Now, $Est\ \frac{1}{N_1} = Est(N_1^{-1})$, let $Est\ N_1 = \hat{N}_1$ (say). But we know that

$$\hat{N}_1 = Est\ (Nq)$$

or

$$\hat{N}_1 = Nq = N(1-\hat{p}).$$

Assume $\hat{N}_1 = N_1 + \epsilon$ where $E(\epsilon) = 0$. We may write $(\hat{N}_1)^{-1} = \{1 + \frac{\epsilon}{N_1}\}^{-1} \cong (N_1)^{-1}\{1 - \frac{\epsilon}{N_1}\}$, neglecting the power of higher than one. Thus,

$$E\left(\frac{1}{N_1}\right) \cong E\left(\frac{1}{N}\right) - \frac{E(\epsilon)}{N_1^2}.$$

Therefore,

$$Est\ \frac{1}{N_1} \cong \frac{1}{\hat{N}_1}\frac{1}{N\hat{q}} \text{ as } \hat{N}_1 = N\hat{q}.$$

Hence,

$$Est\ \frac{N_1 - n_1}{N_1 n_1} \cong \frac{N\hat{q} - n_1}{N\hat{q} n_1} = \frac{N(1-\hat{p}) - n_1}{N(1-\hat{p})n_1}.$$

Putting these values in (16) we obtain

$$\begin{aligned}
\hat{V}(\hat{Y}_{N_1}) &= N\left\{\frac{N(1-\hat{p}) - n_1}{N(1-\hat{p})n_1} s_1^2\right\}\left\{N + (N-1)\hat{p}\left(\frac{n_2-2}{n-2}\right) + \hat{p}\right\} \\
&+ N\bar{y}_1^2\{N\hat{p}^2 - (N-1)\hat{p}\left(\frac{n_2-2}{n-2}\right) - \hat{p}\} \\
&= N\left\{\frac{N(1-\hat{p}) - n_1}{N(1-\hat{p})n_1} s_1^2\right\}\left\{N + (N-1)\hat{p}\left(\frac{n_2-2}{n-2}\right) + \hat{p}\right\} \\
&+ N\hat{p}\bar{y}_1^2\{N\hat{p} - (N-1)\left(\frac{n_2-2}{n-2}\right) - 1\}
\end{aligned} \tag{17}$$

or,

$$\begin{aligned}
\hat{V}(\hat{Y}_{N_1}) = &\frac{N-n+1}{n_1} s_1^2 \left\{N + \frac{(N-1)(n_2-1)(n_2-2)}{(n-1)(n-2)} - \frac{n_2-1}{n-1}\right\} \\
&+ N\bar{y}_1^2\left(\frac{n_2-1}{n-1}\right)\left\{N\left(\frac{n_2-1}{n-1}\right) - \frac{(N-1)(n_2-2)}{n-2} - 1\right\}
\end{aligned}. \tag{18}$$

As

$$(1-\hat{p}) = \frac{n-n_2}{n-1} = \frac{n_1}{n-1}$$

for $n = n_1 + n_2$ Therefore, estimate of the $V(\hat{Y}_{N_1})$ is function of n, n_1, n_2, N, s_1^2, and $\bar{y}_1^2$. These values can be obtained with the help of the samples. Estimate of variance of the estimate of the target population total can be obtained in case of imperfect frame arising due to rare missing units. Estimate of $V(\hat{Y}_{N_1})$ increases as N and n_1 increases but decreases as n increases. If we have $\hat{p} = 1$ so that $n = n_2$ and $n_1 = 0$ then s_1^2, in such case, can not be estimated because $1/n_1$ tends to infinity in the first term of the above equation. This case may arise when there is total imperfection in the frame.

In case of $n_1 = 1$, we have $n_2 = n+1$. In such case $s_1^2 = 0$. Because, s_1^2 can not be determined for one observation. Therefore

$$V(\hat{Y}_{N_1}) = N\bar{y}_1^2 \left\{ N\hat{p}^2 - (N-1)\hat{p}\left(\frac{n_2-2}{n-2}\right) - \hat{p} \right\}.$$

As first term vanishes and $\bar{y}_1^2 = y_1^2$. However, if $\hat{p} = 0$ which may be the case of no imperfection in the frame, then $\hat{V}(\hat{Y}_{N_1})$ is obtained as

$$\hat{V}(\hat{Y}_{N_1}) = \frac{N-n_1}{n_1} s_1^2 N = \frac{N(N-n_1)}{n_1} s_1^2.$$

Since, when $\hat{p} = 0, n_1 = n$ and $s_1^2 = s^2$ therefore

$$V(\hat{Y}_{N_1}) = \frac{N(N-n)}{n} s^2$$

which is obtained in case of perfect frame with S.R.S.W.O.R., this case arises when there is no rare unit in the sample which is non-existent in the target population.

Example: In Chhattisgarh State, the Chhattisgarh Renewable Energy Development Agency (CREDA) had installed 23953 biogas plants by year 2010. During passage of time, it was indicated that there are some plants which become in non-working conditions, annually. The sampler desires to estimate total working plants and total biogas production in the state. Since, number of non-working plants was not known in advance, therefore, the sampling frame is incomplete. The inverse sampling methodology was used to estimate working and non-working plants and total biogas production along with total biogas loss was estimated. with the help of imperfect frame. The non-working plants to be selected were fixed as 18, which were not identified in advance. To get 18 non-working plants, 117 plants were observed. The average production of working biogas plants was found to be 2.8 m^3. Here,

$$n_2 = 18, n_1 = 99, n = 117, N = 23953$$

thus,

$$\hat{p} = 0.146, \hat{q} = 0.854, \bar{y}_1 = 2.8m^3.$$

Therefore, estimated total number of working biogas plants, $N\hat{q} = 20442$ and non-working plants as $N\hat{p} = 3497$. From equation (7), we get total production of biogas in the state as 57276.4 m^3 and the total loss of biogas due to non-performance of plants is found to be 9791.6 m^3.

9. Conclusion

While planning sample survey or census, the existence of sample frame comprising a list of the all the sampling units is pre-requisite. But unfortunately, this situation is hardly achieved in practice and frames are quite often imperfect and incomplete. This may also arise when some of the rare units are missing out- dated or may be out of scope at the time, sampler desire to use. The appropriate and suitable method of estimation is proposed when sampling is done from imperfect frame as elucidated in the formulae explained and illustrated above in section 5 - 8.

Acknowledgements

Thanks are due to a referee for helpful comments that led to a much improved version of the paper.

References

Agarwal, B. and Gupta, P. C. (2008). Estimation from incomplete sampling frame in case of simple random sampling. *Model Assisted Statistics and Applications*, **3**, 113-117.

Agarwal, B. and Singh, S. (2017). Estimation from super-population in case of finite population with incomplete sampling frame. *International Journal of Research and Scientific Innovation*, **4**, 1-4.

Champman, D. G. (1952). Inverse, multiple and sequential sample censuses. *Biometrika*, **8**, 286-288.

Chikkagudar, M. S. (1969). Inverse sampling without replacement. *Australian Journal of Statistics*, **11**, 155-165.

Finney, D. J. (1949). On the method of estimating frequencies. *Biometrika*, **36**, 233-234.

Haldane, J. B. S. (1945). On a method of estimating frequencies. *Biometrika*, **33**, 222-225.

Hansen, M. H., Hurwitz, W. N., and Jabine, T. B. (1963). Use of imperfect tests for probability sampling at the U.S. Bureau of Census. *Bulletin of the International Statistical Institute*, **40**, 497-517.

Hartley, H. O. (1962). *Multiple Frame Surveys*. In Proc. Social Statistics Section Amer. *Statist. Assoc., Annual Meeting, Minneapolis, Minnesota*, 203-206.

Seal, K. C. (1962). Use of outdated frames in large scale sample surveys. *Calcutta Statistical Association Bulletin*, **11**, 68-84.

Singh, N. K. (2020). Frame error in sample survey. *International Research Journal of Agricultural Economics and Statistics*, **11**, 240-244.

Singh, N. K. (2020). Sampling with imperfect frame in large population. *International Research Journal of Agriculture Science and Research*, **10**, 105-112.

Singh, N. K. (2021). Domain studies with imperfect frame in large population. *International Journal of Agricultural Sciences*, **17**, 522-527.

Singh, N. K. (2022). Review On dealing with the problem of imperfect frame in sample survey. *Journal of Emerging Technologies and Innovative Research*, **9(1)**, a690-a699.

Singh, N. K., Kumar, R., and Sehgal, V. K. (2001). Use of incomplete frame on large scale sample survey. *Gujarat Statistical Review*, **28**, 3-10.

Singh, N. K., Sehgal, V. K., and Kumar, R. (1997). Use of incomplete frame for domain studies. *Journal of Indian Statistical Association*, **35**, 71-81.

Singh, R. (1983). On the use of incomplete frame in sample survey. *Biometrical Journal*, **25**, 545-549.

Singh, R. (1989). Method of estimation for sampling from incomplete frames. *Australian Journal of Statistics*, **31**, 269-276.

Szameitat, K. and Schaffer, K. A. (1963). Imperfect frames in statistics and consequences of their use of sampling. *Bulletin of International Statistical Institute*, **40**, 517-538.