Jour. Ind. Soc. Ag. Statistics
49 (Golden Jub. No.), 1996-97 : 203–214

Estimation of Proportions Under Successive Sampling Scheme

Padam Singh and Nishi[1]
Institute for Research in Medical Statistics, New Delhi, India

SUMMARY

A successive sampling scheme for the estimation of proportion is proposed in which in addition to estimating proportion on current occasion it is possible to estimate the proportion of changes in the attribute status. The proposed scheme is developed using successive sampling scheme approach using *srswor* for selection. The efficiency of the proposed sampling scheme over the usual successive sampling scheme has been examined and illustrated under two different situations, one related to Adoption of high yielding variety and another Proportion of households below certain income level. It has been observed that there is considerable gain in efficiency having the proposed sampling as compared to the usual successive sampling scheme. Further, from the proposed scheme it has been possible to estimate the continuation rates and the dropout rates over two periods.

Key words : Proportion, High yielding variety, Poverty line, Successive sampling.

1. Introduction

The problem of estimation of proportions under successive sampling has not been examined as an independent entity and instead has been considered as a special case of estimation of mean. In the present study, a new sampling scheme has been proposed for estimation of proportions following the successive sampling approach. The relevant theory and estimation procedures have been developed and the expressions for the estimators and their variances have been derived. Efficiencies of the proposed schemes have been compared with the usual successive sampling. The use and efficiency of proposed scheme has been illustrated in two different situations (a) Adoption of high yielding variety and (b) Proportion of households below certain income level.

2. *Proposed Scheme*

Consider a population consisting of N units out of which N_1 units possess the attribute and $(N - N_1)$ do not possess the attribute on the first occasion. While on the second occasion from the population of N_1 units, N_{11} possess the attribute on the first as well as on second occasion and N_{12} possess the attribute on the first but do not possess it on the second occasion. Similarly from a population of N_2 units, N_{21} possess the attribute on the second but not on the first occasion and N_{22} do not possess it on the first as well as on second occasion.

The proposed sampling scheme for two occasions consists of the following steps.

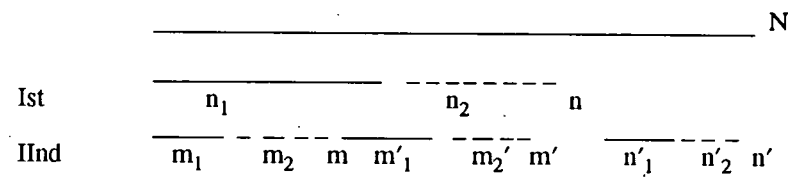
(a) Select a sample of size n by *srswor* from a population of size N and suppose n_1 units out of n possess the attribute. The remaining n_2 units in the sample do not possess the attribute.

(b) From the sample of size n_1 units possessing the attribute on first occasion, a sub-sample of size m is drawn by *srswor* on second occasion. Suppose m_1 units out of m units possess the attribute on the second occasion. Let $m_2 = m - m_1$, be the number of units out of m not possessing the attribute on the second occasion.

(c) From the sample of n_2 units not possessing the attribute a sample of size m′ units is drawn on second occasion. Suppose m_1' units out of m′ possess the attribute. Let $m_2' = m' - m_1'$, be the number of units out of m′ not possessing the attribute on the second occasion.

(d) On the second occasion, in addition, a fresh sample of size n′ is taken from rest of the population (N– n). Suppose n_1' units out of n′ possess the attribute.

Diagramatically, the proposed sampling scheme can be shown as follows :



Here, bold lines indicate units possessing the attribute and broken lines not possessing the attribute.

Notations

Let

$P = N_1 / N$	The proportion of units which possess the given attribute in the population on the first occasion.
$Q = 1 - P$	The proportion of units which do not possess the given attribute in the population on the first occasion.
$p = n_1 / n$	Estimate of proportion who possess the attribute on the first occasion.
$q = n_2 / n$	Estimate of proportion who do not possess the attribute on the first occasion and is such that p+q=1.
$P_1 = N_{11} / N$	The proportion of units possessing the attribute in the population on second occasion from those having attribute on the first occasion.
$p_1 = m_1 / m$	Estimate of P_1.
$P_2 = N_{21} / N$	The proportion of units having attribute in the population on second occasion but not on the first occasion.
$p_2 = m_1' / m'$	Estimate of P_2.
$P_1' = N_1' / N'$	The proportion of units in the population on the second occasion possessing the given attribute.
$p_1' = n_1' / n'$	Estimate of P_1'.
$P_{T2} =$	The proportion of total units possessing the attribute on the second occasion.

(1)

3. *Estimation Procedure*

For second occasion the proposed esimator of $\hat{P}_{T2}$ is as follows

$$\hat{P}_{T2} = a\left[\left(\frac{n_1}{n}\right)\left(\frac{m_1}{m}\right)+\left(\frac{n_2}{n}\right)\left(\frac{m_1'}{m'}\right)\right] + \{1-a\}\left(\frac{n_1'}{n'}\right) \tag{2}$$

$$E(\hat{P}_{T2}) = a\left[E\left(\frac{n_1}{n}\right)\left(\frac{m_1}{m}\right) + E\left(\frac{n_2}{n}\right)\left(\frac{m_1'}{m'}\right)\right] + \{1-a\}E\left(\frac{n_1'}{n'}\right) \tag{3}$$

$$= a\,(P_1 + P_2) + (1-a)P_1'$$

For given 'a' this is an unbiased estimator for P_{T2}. To find the optimum value of 'a' we determine the variance of $\hat{P}_{T2}$.

Variance of $\hat{P}_{T2}$ is given by

$$V(\hat{P}_{T2}) = a^2 V\left[\left(\frac{n_1}{n}\right)\left(\frac{m_1}{m}\right)+\left(\frac{n_2}{n}\right)\left(\frac{m_1'}{m'}\right)\right] + (1-a)^2 V\left(\frac{n_1'}{n'}\right) \quad (4)$$

or

$$V(\hat{P}_{T2}) = a^2 \left\{ V\left(\frac{n_1}{n}\right)\left(\frac{m_1}{m}\right)+ V\left(\frac{n_2}{n}\right)\left(\frac{m_1'}{m'}\right) + 2Cov\left(\frac{n_1}{n}\frac{m_1}{m}, \frac{n_2}{n}\frac{m_1'}{m'}\right)\right\}$$
$$+ (1-a)^2 V\left(\frac{n_1'}{n'}\right) \quad (5)$$

The covariance term in the expression (5) is zero as the two variables $\left(\frac{n_1}{n}\frac{m_1}{m}, \frac{n_2}{n}\frac{m'_1}{m'}\right)$ are independent.

The product formula is given by

$$V(xy) = Ey^2 V(x) + Ex^2 V(y) + V(x) V(y) - 2ExEyCov(x, y)$$
$$- Cov^2(x, y) + Cov(x^2, y^2) \quad (6)$$

For the proposed estimator the covariance terms in product formula will be zero since the sub-sample m drawn on the second occasion from sample n_1, which is a random variable for the first occasion and become fixed on the second occasion. Thus

$$V(\hat{P}_{T2}) = a^2 \left\{ E\left(\frac{m_1}{m}\right)^2 V\left(\frac{n_1}{n}\right)+ E\left(\frac{n_1}{n}\right)^2 V\left(\frac{m_1}{m}\right)+ V\left(\frac{n_1}{n}\right) V\left(\frac{m_1}{m}\right)\right.$$
$$\left. + E\left(\frac{m_1'}{m'}\right)^2 V\left(\frac{n_2}{n}\right)+ E\left(\frac{n_2}{n}\right)^2 V\left(\frac{m_1'}{m'}\right)+ V\left(\frac{n_2}{n}\right) V\left(\frac{m_1'}{m'}\right)\right\}$$
$$+ (1-a)^2 V\left(\frac{n_1'}{n'}\right) \quad (7)$$

Ignoring f.p.c. and substituting in (7), we have

$$V(\hat{P}_{T2}) = a^2 \left\{ \left(P_1{}^2 + \frac{P_1 Q_1}{m} \right) \frac{PQ}{n} + \left(P^2 + \frac{PQ}{n} \right) \frac{P_1 Q_1}{m} + \left(\frac{PQ}{n} \frac{P_1 Q_1}{m} \right) + \left(P_2{}^2 + \frac{P_2 Q_2}{m'} \right) \frac{PQ}{n} + \left(Q^2 + \frac{PQ}{n} \right) \frac{P_2 Q_2}{m'} + \frac{PQ}{n} \frac{P_2 Q_2}{m'} \right\} + (1-a)^2 \frac{PQ}{n'} \tag{8}$$

Above expression is in terms of variance due to follow up component and the variance for the component from an independent fresh sample can be written as

$$V(\hat{P}_{T2}) = a^2 L + (1-a)^2 \frac{PQ}{n'} \tag{9}$$

where

$$L = \frac{PQ}{n} \left(P_1{}^2 + 3 \frac{P_1 Q_1}{m} + P_2{}^2 + 3 \frac{P_2 Q_2}{m'} \right) + P^2 \left(\frac{P_1 Q_1}{m} \right) + Q^2 \left(\frac{P_2 Q_2}{m'} \right) \tag{10}$$

For calculating the optimum value of a, differentiating (9) partially w.r.t. 'a' and equating to zero we have

$$a = \frac{PQ/n'}{L + PQ/n'} \tag{11}$$

In order to know whether the value of 'a' at (11) corresponds to minimum value of V at (9), again differentiate (9) partially w.r.t. 'a' we get

$$\frac{\partial^2 V}{\partial a^2} = 2L + \frac{2PQ}{n} \text{ ; which is positive.} \tag{12}$$

Thus the value of 'a' at (11) corresponds to the minimum value of V at (9). For this value of 'a', variance of the proposed estimator given at (2) for second occasion is

$$V(\hat{P}_{T2}) = \frac{L}{1 + \left(\frac{n'}{PQ} \right) L} \tag{13}$$

or

$$V(\hat{P}_{T2}) = \frac{PQ}{n'} \left\{ 1 - \frac{1}{1 + (n'/PQ)L} \right\} \tag{14}$$

The variance of the usual successive sampling scheme say $V(\bar{y}_2)$ is

$$V(\bar{y}_2) = \frac{PQ}{n}\,\frac{(1-\rho^2\lambda)}{(1-\rho^2\lambda^2)} \tag{15}$$

The value of ρ^2 in terms of proportions is given as

$$\rho^2 = \left\{\frac{\frac{-(PP_1)}{n}}{\sqrt{\frac{(PQ)}{n}\,\frac{(P_1Q_1)}{m}}}\right\}^2 = \left(\frac{mPP_1}{nQQ_1}\right) \tag{16}$$

as $$\mathrm{Cov}(\hat{P}, \hat{P}_1) = \frac{-(PP_1)}{n} \quad \text{and} \quad \lambda = \frac{n'}{n} \tag{17}$$

On substituting the values of ρ^2 and λ given at (16) and (17) in the expression (15) the usual successive sampling scheme variance say $V(\bar{y}_2)$ in terms of proportion reduces to

$$V(\bar{y}_2) = PQ\left(\frac{QQ_1\,n^2 - PP_1\,m\,n'}{QQ_1\,n^2\,n - PP_1\,m\,n'^2}\right) \tag{18}$$

4. *Empirical Comparison*

It is difficult to compare the variance of the proposed scheme with the usual successive sampling scheme mathematically. However, the comparison can be made on the basis of empirical illustrations. Here, the following two applications are presented as an illustration to this problem and efficiency of the proposed sampling scheme for estimation of proportions for two occasions has been compared with the usual successive sampling scheme.

Illustration 1 : Adoption of high yielding variety

A simple random sample of 115 farmers was selected concerning adoption of high yielding variety of Wheat on first occasion. Of these 115 farmers, 58 were adopters and 57 were non-adopters. In a repeat survey on second occasion, a sample of 20 farmers from the first occasion was followed from the adopters group of the first occasion. Of these, proportion of 0.8 persons were found adopting high yielding variety on second occasion as well. Further, a sample of 19 farmers was drawn from the non adopters group and it was found that a proportion of 0.3 of non-adopters on first occasion, started adopting on second occasion. In addition to it a fresh sample of 100 farmers was drawn from the

rest of the population of which 65 were found adopting high yielding variety on second occasion.

The data using the proposed sampling scheme was collected in order to find out the following rates :

(a) Continuation rate of adoption of high yielding variety,

(b) Drop out rate in the adoption of high yielding variety and

(c) Adopters of high yielding for the second occasion.

Now in terms of the notations defined for present study, we have the proportion of adoptors of high yielding on first occasion in situation is $\hat{P} = 58 / 115 \cong 0.50$.

(a) The estimated continuation rate over the two periods for situation is $\hat{P}_1 = m_1 / m = 0.80$ i.e. 80% with s.e. $= \sqrt{p_1 q_1 / m} = 0.089$.

(b) The estimated drop out rate $\hat{Q}_1 = 1 - \hat{P}_1$ for situation is 0.20 i.e. 20% with s.e. $= \sqrt{p_1 q_1 / m} = 0.089$.

(c) The estimated proportion of new cases of adopters of high yielding variety $\hat{P}_2 = m_1' / m'$ for the current second occasion is 0.30 i.e. 30% with s.e. $= \sqrt{p_2 q_2 / m'} = 0.105$.

The estimated proportion of adopters of high yielding variety for the second occasion can be computed as follows :

On the basis of notations defined in the present paper we have

$a = 0.259$, $1 - a = 0.741$ and $L = 0.00647$ thus

$$\hat{P}_{T2} = 0.259\ (0.5*0.80 + 0.5*0.30) + 0.741*0.65$$

or

$$\hat{P}_{T2} = 0.623 \text{ i.e. } 62.3\% \text{ with s.e.} = \sqrt{V(\hat{P}_{T2})} = 0.040$$

Further on comparing the efficiencies of the proposed sampling scheme with the usual sampling scheme it was found that the gain for second occasion over usual sampling scheme for the situation is about 7.2% (Table 1).

Illustration 2 : Proportion of households below certain income level (poverty line)

A simple random sample of 115 persons was selected from the population surveyed from an area of Old Delhi city. Of 115 persons selected from the area, a proportion below annual income of Rs. 6500/- (approximately, the

poverty line as defined by Planning Commission) was found to be 0.35. In a follow-up period, a sample of 15 persons below this income level from the first occasion was followed on the second occasion. Of these, a proportion of 0.6 persons was having income below Rs. 6500/- on the second occasion as well. Of the remaining group of income more than Rs. 6500/- from occasion first a sample of 38 persons was drawn on the second occasion, among this group a proportion of 0.4 persons was found having income below Rs. 6500/- on second occasion. In addition a fresh sample of 95 persons was selected from the rest of the population on the second occasion, of which 0.4 were having income below Rs. 6500/-. It is proposed to use the above data to find out:

(a) Proportion crossing over the poverty line in two periods,

(b) Proportion of persons below poverty line on both the occasions and

(c) The proportion of new households having income less than Rs. 6500/- for the second period.

Now in terms of the notations defined in paper we have

(a) The estimated proportion of persons crossing poverty line

$$\hat{Q}_1 = m_2 / m = 0.4 \text{ i.e. } 40\%$$

(b) The estimated proportion of persons below poverty line on both the occasions, say $\hat{P}_{11}$ given as

$$\hat{P}_{11} = (p\, p_1) = (0.35*0.6) = 0.21 \text{ i.e. } 21\%$$

with s.e. $= \sqrt{V(p_{11})} = 0.052$, where

$$V(p_{11}) = \left\{ \left(P_1^2 + \frac{P_1 Q_1}{m} \right) \frac{PQ}{n} + \left(P^2 + \frac{PQ}{n} \right) \frac{P_1 Q_1}{m} + \left(\frac{PQ}{n} * \frac{P_1 Q_1}{m} \right) \right.$$

$$= 0.00276$$

(c) The estimated proportion of new households having income less than Rs. 6500/- for the second occasion

$$\hat{P}_2 = m_1' / m' = 0.40 \text{ i.e. } 40\% \text{ with s.e.} = \sqrt{p_2 q_2 / m'} = 0.079$$

The estimated proportion of households having income below Rs. 6500/- for the second occasion, can be computed on the basis of notations defined for the proposed estimator in the present study. Thus we have

$$a = 0.29,\ 1 - a = 0.71 \text{ and } L = 0.00618$$

$$\hat{P}_{T2} = 0.29\,(0.35*0.60 + 0.65*0.40) + 0.71*0.4$$

$$= 0.42 \text{ i.e. } 42\%$$

Further the comparison of efficiencies of the proposed sampling scheme with the usual successive sampling scheme under the different situations are given in Table 1.

Table 1. Comparison of variances of the two schemes

Situations	n	m + m′	n′	p	Variances		% Gain
					Proposed scheme	Usual scheme	
Adoption of high yielding variety	115	39	100	0.65	0.00168	0.00181	7.2
Proportion of households below certain income	115	53	95	0.35	0.00179	0.00195	8.2

where

n is the size of the original sample taken on first occasion,

n′ is the size of the fresh sample from (N-n) units on second occasion,

p is the proportion of the units having attribute on the first occasion,

m+m′ is size of total sample taken on second occasion from first occasion.

5. *Cost Function*

Consider the following cost function for a fixed cost say C,

$$C = c\,n + c'\,n' \tag{19}$$

where c is the cost of enumerating the sample n on the first occasion and

c′ is the cost of enumerating the fresh sample n′ on the second occasion.

The size of the fresh sample for different cost ratio c/c′ will be different. Mathematically, it is not easy to show that how much, for different cost ratio (c/c′), the proposed scheme is efficient over the usual successive sampling

scheme. However, with the same illustrations considered, it has been observed that as the ratio of c/c′ decreases, the efficiency of the proposed scheme increases as compared with the usual successive sampling scheme (Tables 3 and 4).

It is observed from Table 2 that as the ratio of the cost of enumerating the sample size drawn on first occasion to the cost of enumerating the sample drawn on the second occasion decreases the size of the fresh sample taken on second occasion increases.

Table 2. Size of the fresh sample for different cost ratios

Situations	n′ for c/c′					
	1.0	0.90	0.75	0.50	0.30	0.10
Adoption of high yielding variety	85	96.5	113.7	142.5	165.5	188.5
Proportion of households below certain income	85	96.5	113.7	142.5	165.5	188.5

Table 3. Comparison of efficiency of the proposed scheme for different cost ratios

Adoption of high yielding variety

For c/c′	Variances		% Gain
	Proposed	Usual scheme	
0.90	0.00172	0.00178	3.4
0.75	0.00152	0.00212	27.6
0.50	0.00128	0.00500	74.4

For decreased cost ratio (c/c′) it has been found that the proposed scheme may be even 74% more efficient than the usual successive sampling scheme.

Table 4. Comparison of efficiency of the proposed scheme for different cost ratios

Proportion of households below certain income level

For c/c′	Variances		% Gain
	Proposed	Usual scheme	
1.00	0.00193	0.00193	–
0.90	0.00176	0.00194	10.2
0.75	0.00157	0.00197	25.4
0.50	0.00132	0.00200	51.4
0.30	0.00117	0.00210	79.4
0.10	0.00101	0.00230	130.0

For decreased cost ratio (c/c′) it has been found that the proposed scheme may be even 130% more efficient than the usual successive sampling scheme.

6. *Conclusions*

Though it is difficult to compare mathematically the variances of the two schemes. However with the empirical procedure, it has been observed that the proposed sampling scheme for two occasions is more efficient than the usual successive sampling variance. To assess the effect of size of follow up sample and the proportion of units drawn from first occasion to second occasion under different illustrations the variance of the proposed scheme for two occasions has been calculated and compared with usual successive sampling variance. It has been observed that the proposed scheme is about 10% more efficient than the usual successive sampling scheme under illustrations considered for the present study. Further, the continuation rates, drop out rates and new cases on the second occasions are also calculated for the current (second) occasion with the proposed scheme which may not be possible using the usual successive sampling scheme. The proposed scheme is more advantageous because of operational convenience, ease of computation and cost effectiveness.

REFERENCES

[1] Cochran, W.G., 1950. *Sampling Techniques*, John Wiley and Sons., New York.

[2] Patterson, H.D., 1950. Sampling on successive occasions with partial replacement of units. *J. Roy. Statist. Soc.*, **B12**, 241-255.

[3] Sukhatme, P.V. and Sukhatme, B.V., 1970. *Sampling Theory of Surveys with Applications*, 2nd ed., Indian Society of Agricultural Statistics, New Delhi and Iowa State University, USA.

[4] Sukhatme, P.V., 1944. Moments and products moments of moment statistics for samples of the finite and infinite population. *Sankhya*, **B**, 363-382.

[5] Yates, F., 1949. *Sampling Methods for Census and Surveys*. Charles Griffin and Co., London.