ESTIMATION OF FREQUENCY DISTRIBUTIONS FOR THE CURRENT OCCASION UNDER SUCCESSIVE SAMPLING APPROACH FOR SOME SELECTED SAMPLING DESIGNS

BY

LOKESH ARORA, *University of Delhi*
AND D. SINGH, *I.A.S.R.I., New Delhi*

(Received : August, 1980)

SUMMARY

The estimation of the frequency distribution for the current occasion using different sampling designs under successive sampling approach has been attempted. In comparing some selected sampling designs for estimation of frequency distribution on the current occasion it has been observed that cluster sampling design with PPS sampling is advantageous in practice.

INTRODUCTION

In the initial stages of the statistical development, whenever, there was a need to study a population, more stress was given to estimate the point parameters, such as, means, totals, ratios and differences. Now there is a change in data requirements from estimates of points parameters to estimates of frequency distributions. This change is very rapid, particularly, in developing countries where the planners and decision-makers are mainly interested in knowing about the structure of the frequency distributions over time and space in planning and to find social and economic development instead of studying the change in averages and ratios. Thus, there is a need to evaluate the sample design for estimating frequency distributions.

Even when the design is meant for estimating point parameters, it is possible to estimate the frequency distributions. The main draw-back is that the estimates of tail-end frequencies are generally very poor and hence not usuable for practical purposes. Mahalanobis (1960) proposed the use of fractiles but it suffers with the deficiencies of the estimates of end class intervals. Dandekar and Rath (1970);

Srinivasan and Bardhan (1974) showed a considerable interest in estimating distribution of house-holds by consumer expenditure in the context of studying poverty in India. For these studies, the users had to use the estimated distribution of consumer expenditure provided by the National Sample Surveys (NSS). Since the NSS estimates are based on general purpose design with emphasis on point parameters, the estimate of tail in wich the users are especially interested, are subject to large sampling errors and hence of limited use for deeper analysis. Murthy (1977) evaluated some selected sample designs for estimating frequency distributions through empirical study.

In the paper, an attempt is made to evaluate some selected sampling designs under successive sampling approach to study the structure of the population by estimating the frequency distributions.

In successive sampling, information collected on the matched portion of the sample is used to improve the estimate of the population character under study. The theory of successive sampling developed by Jessen (1942), Yates (1960), Patterson (1950), Tikkiwal (1951, 56), Eckler (1955) and others for single stage design, was extended to two-stage sampling design by Kathuria (1959), Tikkiwal (1964), Singh (1968), Singh and Kathuria (1969, 71) and Singh and Srivastava (1973).

Estimation of frequency distribution at the recent occasion :

We now consider the estimation of the frequency distributions for the recent occasion by the following selected sampling designs which are of practical interest under successive sampling approach :

(*i*) Direct Sampling of Ultimate Units;

(*ii*) Cluster Sampling of Ultimate Units; and

(*iii*) Two-stage Sampling.

Under these, an estimte of the proportion in the h-*th* class of the frequency distribution ($h=1,2,\ldots,k$), for the current occasion, for each of the sample design, will be obtained and then the efficiencies of the sampling designs will be compared with respect to the error measures $E(\alpha_1)$ and $E(\alpha_2)$, proposed by Murthy (1977).

Let there be N Clusters, further, let M_i ($i=1, 2,\ldots, N$) be the size of i-*th* cluster such that $\sum_{i}^{N} M_i=M$ (the total number of ultimate units in the population). Let m_o be the total number of ultimate units in the sample be expressible as $n\overline{M}$, where n is the number of sample clusters and $\overline{M}=\frac{M}{N}$.

The estimates, under the considered sampling designs along with their respective error measures are discussed below one by one.

(*i*) Direct Sampling of Ultimate Units

Sampling Scheme : Select a sample of m_o ultimate units by SRSWOR from the M ultimate units of the population at first occasion. Obtain the frequency distribution in the desired K classes on the basis of this sample. Retain a sub-sample of $m_o\mu$ $(0 \leqslant \mu \leqslant 1)$ units by SRSWOR from m_o and supplement it with $(m_o - m_o\mu)$ units selected from $(M - m_o)$ units of the population by SRSWOR at second occasion. Obtain the frequency distribution in the desired K classes on the basis of matched sample, unmatched sample, and overall for the first and second occasion both.

The estimate of the proportion, for the current occasion, in h-*th* class of the frequency distribution $(h = 1, 2, \ldots, k)$ is given by—

$$\hat{P}^h_2 = \frac{\mu\lambda\,\sigma^h_{12}}{\sigma_1^{h2}\,(1-\lambda^2\rho^{h2})}\,(p^h_{1\lambda} - p^h_{1\mu}) + \frac{\mu}{(1-\lambda^2\rho^{h2})}\,(p^h_{2\mu} - p^h_{2\lambda}) + p^h_2 \quad \ldots(1)$$

where

$p^h_{1\lambda} =$ the sample proportion of the unmatched units, in *h-th* class of the frequency distribution at first occasion.

$= \dfrac{U^h_1}{m_o\lambda}$, U^h_1 being the number of unmatched sample units in h-*th* class at first occasion.

$p^h_{2\lambda} =$ the sample proportion of the unmatched units in h-*th* class at second occasion.

$= \dfrac{U^h_2}{m_o\lambda}$, U^h_2 being the number of unmatched sample units in h-*th* class at second occasion.

$p^h_{1\mu} =$ the sample proportion of the matched units in h-*th* class at first occasion.

$= \dfrac{m^h_1}{m_o\mu}$, m^h_1 being the number of matched sample units in h-*th* class at first occasion.

$p^h_{2\mu} =$ the sample proportion of the matched units in h-*th* class at second occasion.

$=\frac{m_2^h}{m_0\mu}$, m_2^h being the number of matched sample units in h-*th* class at second occasion.

The variance of the estimate of proportion for the current occasion in h-*th* class of the frequency distribution ($h=1, 2, ..., k$) is given by

$$V(\hat{P}_2^h)=\frac{\sigma_2^{h2}\left[\sigma_1^{h2}\ \sigma_2^{h2}-\lambda\ (\sigma_{12}^h)^2\right]}{m_0\left[\sigma_1^{h2}\ \sigma_2^{h2}-\lambda^2(\sigma_{12}^h)^2\right]} \quad ...(2)$$

where σ_1^{h2} is the population variance of the ultimate units in h-*th* class at first occasion, σ_2^{h2} is the population variance of the ultimate units in h-tb class at second occasion and σ_{12}^h is the population co-variance between the units lying in h-*th* class at both the occasions.

Here $\sigma_1^{h2}=P_1^h(1-P_1^h)$, $\sigma_2^{h2}=P_2^h\ (1-P_2^h)$

and

$$\sigma_{12}^h=(P_{12}^h-P_1^h\ P_2^h),$$

where

$P_1^h=\frac{M_1^h}{M}$ where M_1^h is the number of ultimate units inh-*th* class at first ocassion. $P_2^h\frac{M_2^h}{M}$.

where M_2^h is the number of ultimate units in h-*th* class at second occasion; and

$P_{12}^h=\frac{M_{12}^h}{M}$, where M_{12}^h is the number of ultimate units lying in h-*th* class at both the occasions.

For measuring error, we define on the lines of Murthy (1977):

$$\alpha_1=\sum_h^L(\hat{P}^h-P^h)^2$$

$$\alpha_2=\sum_h^K\frac{(\hat{P}^h-P^h)^2}{P^h}$$

Then, the error measures $E(\alpha_1)$ and $E(\alpha_2)$ are given by—

$$E(\alpha_1)=\sum_{h}^{K}\frac{[P_2^h(1-P_2^h)][P_1^h P_2^h(1-P_1^h)(1-P_2^h-\lambda(P_{12}^h-P_1^h P_2^h)^2]}{m_o\;[P_1^h P_2^h(1-P_1^h)(1-P_2^h)-\lambda^2(P_{12}^h-P_1^h P_2^h)^2]} \quad ...(3)$$

and

$$E(\alpha_2)=\frac{1}{K}\sum_{h}^{K}\frac{[P_2^h(1-P_2^h)].[P_1^h\;P_2^h(1-P_1^h)\;(1-P_2^h)-\lambda(P_{12}^h-P_1^h P_2^h)^2]}{m_o\;P_2^h[P_1^h\;P_2^h\;(-P_1^h)\;(1-P_2^h)-\lambda^2(P_{12}^h-P_1^h\;P_2^h)^2]} \quad ...(4)$$

ii **Cluster Sampling of Ultimate Units :**

Sampling Scheme : Select n clusters by PPSWR from N clusters at first occasion, retain a sub-sample of $n\mu_c$ $(C\leqslant\mu_c\leqslant1)$ clusters from the n clusters by SRSWOR and supplement it with $(n-n\mu_c)$ clusters selected from $(N-n)$ clusters by PPSWR at second occasion. Obtain the frequency distribution in the desired K classes on the basis of the ultimate units in the clusters selected at both the occasions.

The estimate of the proportion for the recent occasion in h-*th* class of the frequency distribution $(h=1, 2, ..., k)$ is given by—

$$\hat{P}_{2c}^h=\frac{\mu_c\,\lambda_c\,\sigma_{12}^h}{\sigma_{1c}^{h2}(1-\lambda_c^2\;\rho_c^{h2})}\left(p_{1\lambda_c}^h-p_{1\mu_e}^h\right)$$

$$+\frac{\mu_c}{(1-\lambda_c^2\;\rho_c^{h2})}\left(p_{2\mu_e}^h-p_{2\lambda_c}^h\right)+p_{2\lambda_c}^h \quad ...(5)$$

where $p_{1\lambda_c}^h$ = the estimate of the proportion of ultimate units of the unmatched clusters, at first occasion in h-*th* class $=\frac{1}{n\lambda_e}\sum_{i}^{n\lambda_c}\frac{m_{1c}^h}{p_iM}$ where M_{1i}^h being the number of ultimate units of i-*th* unmatched cluster, at first occasion, in h-*th* class and p_i is the probability of selecting i-*th* cluster.

$p_{2\lambda_c}^h$ = the estimate of the proportion of the ultimate units of the unmatched clusters, at second occasion in h-*th* class $=\frac{1}{n\lambda_c}\sum_{i}^{n\lambda_c}\frac{M_{2i}^h}{p_iM}$ where M_{2i}^h being the number of ultimate units of unmatched clusters at second occasion, in h-*th* class.

$p^h_{1\mu_c}$ = the estimate of the proportion of ultimate units of the matched clusters at first occasion, in h-*th* class$=\frac{1}{n\mu_c}\sum_{i}^{r\mu_c}\frac{M^h_{1i}}{p_iM}$, where M^h_{1i} being the number of ultimate units of i-*th* matched cluster, at first occasion, in h-*th* class.

$p^h_{2\mu_c}$ = the estimate of the proportion of ultimate units of the matched clusters at second occasion, in h-*th* class$=\frac{1}{n\mu_c}\sum^{u\mu_c}\frac{M^h_{2i}}{p_iM}$, where M^h_{2i} being the number of ultimate units of the *i-th* matched cluster, at second occasion, in *h-th* class.

Here

$$\sigma^{h2}_{1c}=\left[\sum_{i}^{N}\left(\frac{M_i}{M}\right)^2\frac{(P^h_{1i})^2}{p_i}-(P^h_1)^2\right]$$

$$\sigma^{h2}_{2c}=\left[\sum_{i}^{N}\left(\frac{M_i}{M}\right)^2\frac{(P^h_{2i})^2}{p_i}-(P^h_2)^2\right]$$

and

$$\sigma^h_{12}=\left[\sum_{i}^{N}\left(\frac{M_i}{M}\right)^2\frac{P^h_{1i}\;P^h_{2i}}{p_i}-P^h_1\;P^h_2\right]$$

The variance of the estimate $\hat{P}^h_{2c}$ is given by

$$V(\hat{P}^h_{2c})=\frac{\sigma^{h2}_{1c}\;[\sigma^{h2}_{2c}\,\sigma^{h2}_{2c}-\lambda_c(\sigma^h_{12c})^2]}{n[\sigma^{h2}_{1c}\;\;\sigma^{h2}_{2c}-\lambda^2_c\;(\sigma^h_{12c})^2]}\qquad\ldots(6)$$

The error measure $E(\alpha_1)$ is given by

$$E(\alpha_1)=\sum_{h}^{k}\frac{\sigma^{h2}_{2c}\;[\sigma^{h2}_{1c}\;\;\sigma^{h2}_{2c}-\lambda_c(\sigma^h_{12c})^2]}{n[\sigma^{h2}_{1c}\;\;\sigma^{h2}_{2c}-\lambda^2_c\;(\sigma^h_{12c})^2]}\qquad\ldots(7)$$

and the error measure $E(\alpha_2)$ is given by

$$E(\alpha_2)=\frac{1}{k}\sum_{h}^{k}\frac{\sigma^{h2}_{2c}\;[\sigma^{h2}_{1c}\;\;\sigma^{h2}_{2c}-\lambda_c(\sigma^h_{12c})^2]}{n\,P^l_2\;\;[\sigma^{h2}_{1c}\;\;\sigma^{h2}_{2c}-\lambda^2_c\;(\sigma^h_{12c})^2]}\qquad\ldots(8)$$

(*iii*) **Two-Stage Sampling :**

Sampling Scheme. Select a sample of n primary stage units (p.s.u.) from N p.s.u.'s by PPSWR at first occasion. Retain a sub-sample of size $n\mu_t(0\leqslant\mu_t\leqslant 1)$ from n p.s.u.'s by SRSWOR and supplement it with $(N-n\mu_t)$ p.s.u's selected by PPSWR from $(N-n)$ p.s.u's at the second occasion. Then select a sub-sample of second stage units (s.s.u's) of size m from each of the selected cluster at both the occasions. Find the frequency distribution of s.s.u's of the matched and unmatched clusters at both the occasions in the desired K classes of the frequency distribution.

The estimate of the proportion for the recent occasion in *h-th* class of the frequency distribution ($h=1, 2, ..., k$) is given by

$$\hat{P}_{2t} = \frac{\mu_t\lambda_t\sigma^h_{12t}}{\sigma^{h2}_{1t}\,(1-\lambda^2_t \quad \rho^{h2}_t)}\,(p^h_{1\lambda_t}-p^h_{1\mu_t}) + \frac{\mu_t}{(1-\lambda^2_t\;/\rho^{h2}_t)}\,(p^h_{2\mu_t}-p^h_{2\lambda_t}) + p^h_{2\lambda_t} \qquad ...(9)$$

where

$p^h_{1\lambda_t}$ = the estimate of the proportion of ultimate units of the unmatched p.s.u's at first occasion in *h-th* class

$$= \frac{1}{n\lambda_t}\sum_i^{n\lambda_t}\frac{M_i}{M}\,\frac{p^h_{1i}}{p_i}, \quad \text{where} \quad p^h_{1i} = \frac{U^h_{1i}}{U_{1i}},$$

U_{1i} being the number of s.s.u's selected from *i-th* unmatched p.s.u at first occasion and U^h_{1i} is the number of selected s.s.u's from the *i-th* unmatched p.s.u's at first occasion in *h-th* class.

$p^h_{2\lambda_t}$ = the estimate of the proportion of the ultimate units of the unmatehed p.s.u's at the second occasion in *h-th* class

$$= \frac{1}{n\lambda_t}\sum_i^{n\lambda t}\frac{M_i}{M}\,\frac{p^h_{2i}}{p_i}, \quad \text{where} \quad p^h_{2i} = \frac{U^h_{2i}}{U_{2i}},$$

U_{2i} being the number of ultimate units selected from the *i-th* unmatched p.s.u's on the second occasion and U^h_{2i} is the number of selected s.s.u's from the *i-th* unmatched p.s.u's at second occasion in the *h-th* class.

$p^h_{1\mu_t}$ = the estimate of the proportion of the ultimate units of the matched p.s.u's at first occasion in *h-th* class

$$= \frac{1}{n\mu_t} \sum_i^{n\mu_t} \frac{M_i}{M} \frac{p^h_{1i}}{p_i}, \quad \text{where} \quad p^h_{1i} = \frac{m^h_{1i}}{m_{1i}},$$

m_{1i} being the number of s.s.u's selected from the *i-th* matched p.s.u's at the first occasion and m^h_{1i} is the number of s.s.u's selected from *i-th* p.s.u's at first occasion in *h-th* class.

$p^h_{2\mu_t}$ = the estimate of the proportion of ultimate units of the matched p.s.u's at the second occasion in *h-th* class :

$$= \frac{1}{n\mu_t} \sum_i^{n\mu_t} \frac{M_i}{M} \frac{p^h_{2i}}{p_i}, \quad \text{where} \quad p^h_{2i} = \frac{m^h_{2i}}{m_{2i}},$$

m_{2i} being the number of ultimate units selected from the *i-th* matched p.s.u's at second occasion and m^h_{2i} is the number of ultimate units selected from *i-th* p.s.u's at second occasion lying in *h-th* class.

Here

$$\sigma^{h2}_{1t} = \left[\sum_i^N \left(\frac{M_i}{M} \right)^2 \frac{(P^h_{1i})^2}{p_i} - (P_1^h)^2 \right]$$

$$+ \left[\sum_i^N \left(\frac{M_i}{M} \right)^2 \frac{(M_i - m)\, P^h_{1i}\, (1 - P^h_{1i})}{p_i\, (M_i - 1)\, m} \right]$$

$$\sigma^{h2}_{2t} = \left[\sum_i^N \left(\frac{M_i}{M} \right)^2 \frac{(P^h_{2i})}{p_i} - (P_2^h)^2 \right]$$

$$+ \left[\sum_i^N \left(\frac{M_i}{M} \right)^2 \frac{(M_i - m)\, P^h_{2i} (1 - P^h_{2i})}{p_i\, (M_i - 1)\, m} \right]$$

and

$$\sigma^h_{12t} = \left[\sum_i^N \left(\frac{M_i}{M} \right)^2 \frac{P^h_{1i}\; P^h_{2i}}{p_i} - P_1^h\, P_2^h \right]$$

$$+ \left[\sum_i^N \left(\frac{M_i}{M} \right)^2 \frac{(M_i - m)\, (P^h_{12i} - P^h_{1i}\, P^h_{2i})}{p_i\, (M_i - 1)\, m} \right]$$

The variance of the estimate $\hat{P}^h_{2t}$ is given by

$$V\left(\hat{P}^h_{2t}\right)=\frac{\sigma^{h2}_{2t}\ [\sigma^{h2}_{1t}\ \sigma^{h2}_{2t}\ -\lambda_t\ (\sigma^{h}_{12t})^2]}{n\,[\sigma^{h2}_{1t}\ \sigma^{h2}_{2t}\ -\lambda^2_t\ (\sigma^{h}_{12t})^2]} \qquad ...(10)$$

The error measure $E(\alpha_1)$ is given by

$$E(\alpha_1)=\sum_{h}^{K}\frac{\sigma^{h2}_{2t}\ [\sigma^{h2}_{1t}\ \sigma^{h2}_{2t}\ -\lambda_t\ (\sigma^{h}_{12t})^2]}{n\,[\sigma^{h2}_{1t}\ \sigma^{h2}_{2t}\ -\lambda^2_t\ (\sigma^{h}_{12t})^2]} \qquad ...(11)$$

The error measure $E(\alpha_2)$ is given by

$$E(\alpha_2)=\frac{1}{K}\sum_{h}^{K}\frac{\sigma^{h2}_{2t}\ [\sigma^{h2}_{1t}\ \sigma^{h2}_{2t}\ -\lambda_t\ (\sigma^{h}_{12t})^2]}{n\ P^h_2\ [\sigma^{h2}_{1t}\ \sigma^{h2}_{2t}\ -\lambda^2_t\ (\sigma^{h}_{12t})^2]} \qquad ...(12)$$

It may be noted that the estimates of proportions contain unknown parameters in (1), (5) and (9) may, however, be estimated by the sample values. Further the estimate of the variance and also the error measures can also be obtained by substituting the sample estimates for different components appearing in the expressions.

Some empirical studies

The formulae for the two error measures $E(\alpha_1)$ and $E(\alpha_2)$, in case of the sampling designs, considered here are in a complicated form, therefore, an algebraic comparison of the sampling designs is not only difficult but also futile in practice. Hence, the empirical studies have been made for comparision,though this approach may not lead to general conclusions applicable universally.

For the comparison of different sampling designs, the following two sets of data were used :

(*a*) Data relating to area under bajra for 36 villages of Delhi for the years 1976-77 and 1977-78, collected under High Yielding Variety Survey Programme at Indian Agricultural Statistics Research Institute (IASRI).

(*b*) Data relating to area under wheat for 36 villages of Delhi for the years 1976-77 and 1977-78, collected under High Yielding Variety Survey at IASRI.

From the values of the error measures $E(\alpha_1)$ and $E(\alpha_2)$, for different values of μ and n considered, following observations may be made. It is seen that the value of μ, for which the error measure $E(\alpha_1)$ is minimum, remains the same (for the error measure $E(\alpha_2)$ also) for entire range of the sample size. The optimum value of μ was further seen to be nearly of the same order as in estimating the mean of the population. A consistent trend has been observed for all the three sampling designs and also for both the populations.

It may also be mentioned that the differences between the minimum values of $E(\alpha_1)$ and the maximum values of $E(\alpha_1)$ in the range of the values of μ considered, for a given n, corresponding to the direct sampling of ultimate units, cluster sampling [Equal probability sampling (E.P.S.) and probability proportional to size sampling (P.P.S)] and two-stage sampling (E.P.S. and P.P.S.) were of the order (expressed in percentages and averaged over all n) 4%, (13% and 11%) and (·16% and ·56%) for population A. For population B, the differences between the minimum and the maximum values of $E(\alpha_1)$, in the above order, were noted to be 14%, (25% and 22%) and (·11% and ·09%) respectively.

Similarly the differences between the minimum values of $E(\alpha_2)$ and the maximum values of $E(\alpha_2)$, in the range of μ considered, for the Direct Sampling of Ultimate Units, Cluster Sampling (E.P.S. and P.P.S) and Two-Stage Sampling (E.P.S. and P.P.S.) were observed as 3%, (11% and 7%) and (·12% and ·76%) respectively for population A. For population B, these differences, in the above order, abserved to be 13%, (23% and 21%) and (·07% and ·07%) respectively.

It is well known that Cluster Sampling and Two-Stage Sampling are potentially better than the Direct Sampling of Ultimate Units when cost aspect and operational convenience are the main considerations. In Cluster and Two-Stage Sampling designs, there is a choice of either using E.P.S. or P.P.S. sampling. The values of two error measures for Cluster Sampling and Two-Stage Sampling designs under E.P.S. and P.P.S. sampling are presented below for different sample size corresponding to that value of μ which gives the minimum error measures in entire range considered. The results of the Direct Sampling of units are also included for reference purposes. These results are being presented in Tables (A.1, A.2) and (B.1, B.2) for population A and B respectively.

TABLE A1

The values of the error measure E(α_1), for the considered sampling designs, for different samples sizes, for population A.

S. No.	*Direct sampling of ultimate units*		*Cluster sampling*				*Two-Stage Sampling*			
	m_0	*Error measure*	*n,*	$\overline{M}$	*E.P.*	*P.P.S.*	*n*	*m*	*E.P.*	*P.P.S.*
1	30	.0202	3	10	.0544	.0490	5	6	.0585	.0519
2	60	.0101	6	10	.0272	.0245	10	6	.0293	.0259
3	90	.0067	9	10	.0181	.0163	15	6	.0195	.0173
4	120	.0050	12	10	.0136	.0123	20	6	.0146	.0130
5	150	.0040	15	10	.0109	.0098	25	6	.0117	.0104
6	180	.0034	18	10	.0091	.0082	30	6	.0098	.0086
7	360	.0017	36	10	.0045	.0041	36	6	.0081	.0072

TABLE A2

The values of the error measure E(α_2), for the considered sampling designs, for different sample sizes, for population A.

S. No.	*Direct sampling of ultimate units*		*Cluster sampling*				*Two-Stage sampling*			
	m_0	*Error measure*	*n*	$\overline{M}$	*E.P.*	*P.P.S.*	*n*	*m*	*E.P.*	*P.P.S.*
1	30	.0216	3	10	.0537	.0482	5	6	.0609	.0538
2	60	.0108	6	10	.0269	.0241	10	6	.0304	.0269
3	90	.0072	9	10	.0179	.0161	15	6	.0203	.0179
4	120	.0054	12	10	0.134	.0120	20	6	.0152	.0135
5	150	.0043	15	10	.0108	.0096	25	6	.0123	.0108
6	180	.0036	18	10	.0089	.0080	30	6	.0101	.0090
7	360	.0018	36	10	.0044	.0040	36	6	.0085	.0075

TABLE B1

The values of the error measure E(α_2), for the considered sampling designs, for different sample sizes, for population B.

S. No.	*Direct sampling of ultimate units*		*Cluster sampling*				*Two-Stage Sampling*			
	m_0	*Error measure*	*n,*	$\overline{M}$	*E.P.*	*P.P.S.*	*n*	*m*	*E.P.*	*P.P.S.*
1	30	.0187	3	10	.0382	.0345	5	6	.0537	.0473
2	60	.0093	6	10	.0191	.0173	10	6	.0268	.0236
3	90	.0062	9	10	.0127	.0115	15	6	.0179	.0157
4	120	.0047	12	10	.0096	.0086	20	6	.0134	.0118
5	150	.0037	15	10	.0076	.0069	25	6	.0107	.0094
6	180	.0031	18	10	.0064	.0058	30	6	.0089	.0079
7	360	.0016	36	10	.0032	.0029	36	6	.0075	.0066

TABLE B2

The values of the error measure E(α_3), for the considered sampling designs, for different sample sizes, for population B.

S. No.	*Direct sampling of ultimate units*		*Cluster samplng*				*Two-Stage sampiing*			
	m_0	*Error measure*	*n,*	$\overline{M}$	*E.P.*	*P.P,S.*	*n*	*m*	*E.P.*	*P.P.S.*
1	30	.0194	3	10	.0379	.0346	5	6	.0547	.0484
2	60	.0097	6	10	.0189	.0173	10	6	.0274	.0242
3	90	.0065	9	10	.0126	.0115	15	6	.0182	.0161
4	120	.0049	12	10	.0095	.0086	20	6	.0137	.0121
5	150	.0039	15	10	.0076	.0069	25	6	.0109	.0097
6	180	.0032	18	10	.0063	.0058	30	6	.0091	.0081
7	360	.0016	36	10	.0032	.0029	36	6	.0076	.0067

It can be seen from the above Tables that there is gain in efficiency using P.P.S. Sampling over E.S.S. for both Cluster as well as Two-Stage Sampling designs. Of the two, the gain in efficiency using P.P.S. Sampling in Cluster Sampling design is observed to be of higher order as compared to the corresponding gain in the Two-Stage Sampling design. Further, this gain becomes smaller with increase in the sample size for both the sampling designs. Thus, keeping in view all the factors, such as, operational convenience, cost aspect and the magnitude of error measures, the use of Cluster Sampling with P.P.S. Sampling can be adopted with greater advantages.

ACKNOWLEDGEMENT

The authors are thankful to the referee for his valuable comments in improving the paper.

REFERENCES

Dandekar, V.M. and Rath, N. (1971) : *Poverty in India*, Indian School of Political Economy, Pune.

Eckler, A.R. (1955) *Rotation Sampling*, A.M.S., Vol. **26**, 664-685.

Jessen, R.J. (1942) : 'Statistical investigation of a Sample Survey for obtaining farm facts', *Iowa Agri. Expt. Station Res. Bull.* No, 304, Ames Ia.

Kathuria, O.P. (1959) : 'Some aspects of succhssive Sampling in multi-stage desings', Thesis submitted for award of I.C.A.R., Diploma (unpublished).

Mahalanobis, P.C. (1960) : "A method of fractile graphical analysis", *Econometrica*, 28, 2. 325-351, reprinted in *Sankhya* (1961), 23A, 41-64.

Murthy, M.N. : Presented a paper on 'Use of empirical studies in evaluating sample designs for estimating frequency distribution', at the 41st Session of the International Statistical Institute, New Delhi (5-15 December, 1977).

Patterson, H.D, (1950) : 'Sampling on Successive Occasions with partial replacement of Units', *J.R.S.S., Series B.* Vol. **12**, 241-255.

Singh, D. (1968) : 'Estimation in Successive sampling using a multi-stage design', *Jour. Amer. Stat. Ass.*, Vol. 63, 99-112.

Singh, D. and Kathuria, O.P., (1969) : "On Two-stage successive sampling' *Aust. Jr. Stat.*, Vol. **11**, 59-66.

Singh, D. and Kathuria, O.P. (1971) : 'Comparison of estimates in two-stage Sampling on successive occasions', *Jour. Indi Soc. Agri. Stat.*, Col. **23**, 31-51.

Singh, D. and Kathuria, O.P. (1971) : 'Relative efficiencies of some alternative replacement procedures in two-stage sampling on successive occason', *Jour. Indian Soc. Agri. Stat.*, Vol. 23 (Silver Jubilee Number), 101-114.

Singh, D. and Srivastava, A.K. (1973) : 'Use of auxiliary information in two-stage successive sampling'. *Jour. Ind. Soc. Agri. Stat.*, Vol. 25—I, 101-114.

Srinivasan, T.N. and Bardhan, P.K. (1974) : *Poverty and income distribution in India*, Statistical Publishing Society, Calcutta.

Tikkiwal. B.D. (1956) : 'A further contribution to the theory of univariate sampling on successive occasions', *Jr. Ind. Soc. Agri. Stat.*, Vol. 8, 84-90.

Tikkiwal, B.D. (1964) : 'A note on two-stage sampling on successive occasion', *Sankhya* Series A. Vol. 26, 97-100.

Yates, F. (1960) : *Sampling Methods for Censuses and Surveys*, Charles Griffin and Co., London.