A NOTE ON TWO STAGE SUCCESSIVE SAMPLING

A. K. SRIVASTAVA AND SHIVTAR SINGH
Institute of Agricultural Research Statistics, New Delhi

(Received in September, 1970 ; Accepted in January, 1974)

1. INTRODUCTION

Sampling on successive occasions essentially consists of partially retaining a fraction of the sample selected on the previous occasions, supplemented by fresh units on each occasion and utilising the entire information from the matched units to improve the estimates on later occasions. Repeat surveys are indispensible if one wants to study the changes occurring in the population character from occasion to occasion. The theory has been developed by Jessen (1942), Patterson (1950), Eckler (1955), Tikkiwal (1951, 55, 56), etc. The theory of successive sampling in case of multistage designs has been considered by Tikkiwal (1964, 65), Singh (1968), Singh and Kathuria (1969) and Abraham, Khosla and Kathuria (1969).

In case of two-stage designs the partial retention of primaries and secondaries offers many alternatives. A general case can be considered as retaining a fraction p of primary stage units (psu's) and selecting afresh a fraction q of psu's $(p+q=1)$ and from each retained psu retaining only a fraction r of second stage units ((ssu's) and selecting afresh a fraction s of the ssu's such that $r+s=1$.

2. SAMPLING ON h OCCASIONS

Consider a population consisting of N psu's each containing M ssu's. On the first occasion take a simple random sample (s. r. s.) of n psu's and select a s.r.s. of m ssu's from each of the selected psu's, selection being without replacement at both the stages. For convenience the number of psu's and ssu's selected on each occasion is taken to be constant in the present paper. On the second occasion retain from the first occasion a subsample of size *np* of psu's and select afresh *nq* psu's. In each of the *np* psu's retained, select *mr* ssu's and retain them on the second occasion and select afresh *ms* ssu's. In each of the *nq* psu's, select *m* ssu's following the selection procedure as in the first occasion. The psu's retained during the second occasion will remain fixed for the subsequent occasions but the remaining *nq* psu's will be selected afresh on each occasion. Also from each of the *np* psu's, *mr* ssu's retained on the second occasion will be retained on the subsequent occasions whereas *ms* ssu's will be selected afresh on each occasion.

Let the character under study be X. The following notations are used in the present context:

$\overline{X}_t$: population mean per ssu on the t-th occasion.

$\bar{x}'_t$: mean per ssu on the t-th occasion based on *npmr* units which are common to all the occasions.

$\bar{x}''_t$: mean per ssu on the t-th occasion based on *npms* fresh units taken from common psu's on all the occasions.

$\bar{x}'''_t$: mean per ssu on the t-th occasion based on *nqm* fresh units on the t-th occasion.

The following result due to Tikkiwal (1965, Cor. 2·1, p. 126) will be used :

E_t is the best linear unbiased estimate (BLUE) of $\overline{X_t}$ based on observations up to and including h-th occasion if and only if cov $(E_t, x_{ijk}) = \lambda_i$ for all i, j and k where x_{ijk} is the value of the k^{th} ssu in the j^{th} psu observed on the i-th occasion.

It follows from this result that in order that an estimate E_h may be best linear unbiased estimate of $\overline{X_h}$, it must satisfy the following:

$$\text{Cov}\,(E_h, \bar{x}'_t) = \text{Cov}\,(E_h, \bar{x}''_t) = \text{Cov}\,(E_h, \bar{x}'''_t) \qquad \ldots(1)$$

for all t $(t=1, 2, 3, \ldots, h)$.

An unbiased linear estimate of $\overline{X_h}$ may be given by

$$E_h = \sum_{i=1}^{h} a_{2i-1}\,(\bar{x}'_i - \bar{x}'''_i) + \sum_{i=1}^{h} a_{2i}\,(\bar{x}''_i - \bar{x}'''_i) + \bar{x}'''_h \qquad \ldots(2)$$

By Lemma 2·4 (Tikkiwal, 1965), the variance of E_h, after ignoring terms $0(1/M)$, is seen to be given by

$$V(E_h) = \sum_{i=1}^{2h} \sum_{j=1}^{2h} a_i\, a_j\, \gamma_{ij} + [1 - 2(a_{2h-1} + a_{2h})]\, V(\bar{x}'''_h) \qquad \ldots(3)$$

where

$$\gamma_{ij} = \begin{cases} \left(1/(npq)\right)\left[Sb^2_{i''} + \dfrac{Sw^2_{i''}(q+pr)}{mr}\right]; \ i \text{ is odd} \\ \left(1/(npq)\right)\left[Sb^2_{i'} + \dfrac{Sw^2_{i'}(q+ps)}{ms}\right]; \ i \text{ is even} \end{cases} \text{ when } i=j$$

$$\gamma_{ij} = \begin{cases} \left(1/(npq)\right)\left[Sb^2_{i''} + p\left(\dfrac{Sw^2_{i''}}{m}\right)\right]; \quad i \text{ is odd}; \ j-i=1 \\ \rho'_{i''j''}\dfrac{Sb_{i''}\,Sb_{j''}}{np} + \rho''_{i''j''}\dfrac{Sw_{i''}\,Sw_{j''}}{npmr}; \quad i \text{ and } j \text{ both are odd} \\ \rho'_{i''j'}\dfrac{Sb_{i''}\,Sb_{j'}}{np}; \ i \text{ is odd}, \ j \text{ is even}, j-i>1 \\ \rho'_{i'j''}\dfrac{Sb_{i'}\,Sb_{j''}}{np}; \ i \text{ is even}, \ j \text{ is odd} \\ \rho'_{i'j'}\dfrac{Sb_{i'}\,Sb_{j'}}{np}; \ i \text{ is even}, j \text{ is even} \end{cases} \text{ when } j>i$$

$$i'=i/2, \quad i''=(i+1)/2, \quad j'=j/2, \quad j''=(j+1)/2$$

and

$$Sb^2_t = \left(1/(N-1)\right)\sum_{i=1}^{N}\left(\bar{X}_{ti}-\bar{X}_t\right)^2$$

$$Sw^2_t = \left(1/(M-1)N\right)\sum_{i=1}^{N}\sum_{j=1}^{M}(X_{tij}-\bar{X}_{ti})^2$$

$$\rho'_{tt'}\,Sb_tSb_{t'} = \left(1/(N-1)\right)\sum_{i=1}^{N}(\bar{X}_{ti}-\bar{X}_t\ (\bar{X}_{t'i}-\bar{X}_{t'})$$

$$\rho''_{tt'}\,Sw_t\,S_{wt'} = \left(1/(N(M-1))\right)\sum_{i=1}^{N}\sum_{j=1}^{M}(\overline{X}_{tij}-\overline{X}_{ti})(\overline{X}_{t'ij}-\overline{X}_{t'i})$$

X_{tij}= the observation at the t-th occasion on j-th ssu in the i th psu.

$$\overline{X_{ti}} = (1/M) \sum_{j=1}^{M} X_{tij}.$$

It is evident that $\gamma_{ij} = \gamma_{ji}$.

The optimum values of a_i's, which will minimize $V(E_h)$ are obtained from the following normal equations.

$$PA=B \qquad \ldots(4)$$

where $P = ((\gamma_{ij}))$ is a $2h \times 2h$ matrix of coefficients,

$A = (a_1, a_2, a_3, \ldots\ldots, a_{2h})'$

and

$B = (0, 0, 0, \ldots, V(\bar{x}'''_h), V(\bar{x}'''_h))'$.

Now, if the matrix P is non singular, the coefficients a_i's are determined by $A = P^{-1}B$

or $a_i = (\gamma^{2h-1, i} + \gamma^{2h, i}) V(\bar{x}'''h)/\Delta'$, $(i=1, 2, \ldots, 2h)$

where Δ' is the determinant of the matrix P, γ^{ij} is the cofactor of γ_{ji} in Δ'. With these optimum values of a_i's it can easily be seen that the estimate E_h satisfies the condition (1), for being best linear unbiased estimate of $\overline{X}_h$.

Variance of the estimate E_h is

$$V(E_h) = \text{Cov}(E_h, \bar{x}'''_h) = (1 - a_{2h-1} - a_{2h}) V(\bar{x}'''_h)$$

$$= \frac{a_h}{nq}\left[1 - \frac{\gamma^{2h-1, 2h-1} + 2\gamma^{2h-1, 2h} + \gamma^{2h, 2h}}{\Delta'}\right]$$

where $a_h = Sb^2_h + Sw^2_h/m$

When there are only two occasions, the estimate of the mean on the second occasion is given by

$$E_2 = a_1(\bar{x}_1' - \bar{x}'''_1) + a_2(\bar{x}''_1 - \bar{x}'''_1) + a_3(\bar{x}'_2 - \bar{x}'''_2)$$
$$+ a_4(\bar{x}''_2 - \bar{x}'''_2) + \bar{x}''_2$$

under the assumptions

$Sb^2_1 = Sb^2_2 = Sb^2$; $Sw^2_1 = Sw^2_2 = Sw^2$; $\alpha_t = \alpha$ for $t = 1, 2$
$\rho'_{12} = \rho'$ and $\rho''_{12} = \rho''$ the coefficients are,

$$a_1 = -\frac{p\alpha r}{\Delta''}\left[(q\rho' + \rho'' s) Sb^2 + \rho''(q + ps)\frac{Sw^2}{m}\right]$$

$$a_2 = \frac{p\alpha s}{\Delta''}\left[(r\rho'' - q\rho' + \rho'\rho''^2 qs) Sb^2 + pr\rho'' \frac{Sw^2}{m}\right]$$

$$a_3 = \frac{p\alpha r}{\Delta''}\left[(1 + \rho'\rho'' qs) Sb^2 + \frac{Sw^2}{m}\right]$$

$$a_4=\frac{p\alpha s}{\Delta''}\Big[(1-\rho'\rho''qr-\rho''^2 s)\,Sb^2 + [1-(q+ps)\rho''^2]\frac{Sw^2}{m}\Big]$$

where

$$\Delta''=\Big[(1+\rho'\rho''qs)\,Sb^2+\frac{Sw^2}{m}\Big]^2 - \Big[(q\rho'+\rho''s)\,Sb^2+\rho''(q+ps)\frac{Sw^2}{m}\Big]^2$$

and the variance is given by

$$V(E_2)=(\alpha/nq)(1-a_3-a_4)$$

$$=\frac{\alpha}{nq}\left[[1-p\alpha\,\frac{Sb^2(1-\rho''^2s^2)+\frac{Sw^2}{m}(1-(q+ps)s\rho''^2)}{\Delta''}\right]$$

It can easily be seen that when $p=1$, $q=0$ and when $r=1$, $s=0$, the estimator and its variance reduce to the corresponding cases discussed by Singh and Kathuria (1969).

An estimator [Abraham, Khosla and Kathuria (1969)] of $\overline{X}_2$, by first considering np primary sampling units which are common to two occasions and then utilising the information contained in unmatched primaries by proper weighing, is given as

$$\bar{x}_w=K\,\bar{x}_{2c}+(1-K)\bar{x}'''_2+K'(\bar{x}_1-\bar{x}_1''')$$

where

$$\bar{x}_{2c}=k[\bar{x}'_2+\rho''(\bar{x}_1-\bar{x}'_1)]+(1-k)\bar{x}''_2$$

$$\bar{x}_1=r\bar{x}'_1+s\bar{x}''_1.$$

The optimum values of k, K and K' are

$$k=r/(1-\rho''^2s^2),\qquad K=p\alpha^2/\Delta$$

$$K'=(-qK/\alpha)\,[\rho'Sb^2+\rho''r/(1-\rho''^2s^2)(Sw^2/m)]$$

where

$$\Delta=\alpha\Big[Sb^2+\frac{1-(q+ps)\rho''^2s}{(1-\rho''^2s^2)}\,\frac{Sw^2}{m}\Big]-q^2\Big(\rho'Sb^2+\frac{\gamma\rho''}{1-\rho''^2s^2}\,\frac{Sw^2}{m}\Big)^2$$

The minimum variance of $\bar{x}_w$, is given by

$$V(\bar{x}_w)=\frac{\alpha}{n}\Big[\alpha\Big(Sb^2+\frac{1-\rho''^2s}{1-\rho''^2s^2}\,\frac{Sw^2}{m}\Big) -q\ \rho'Sb^2+\Big(\frac{r\rho''}{1-\rho''^2s^2}\,\frac{Sw^2}{m}\Big)^2\Big]$$

It has been concluded in the paper referred to, by an empirical study, that the estimator $\bar{x}_w$ is equal or more efficient than the estimator E_2 for different values of ρ', ρ'', p, r and $\varphi\left(=\frac{Sw^2}{Sb^2}\right)$. However, it can easily be seen that $\bar{x}_w$ does not satisfy the necessary and sufficient conditions (1) for being minimum variance unbiased linear estimate, for,

$$\text{Cov}(\bar{x}_w, \bar{x}'''_2) = \frac{\alpha}{n\triangle}\left[\alpha\left(Sb^2 + \frac{1-\rho''^2 s}{1-\rho''^2 s^2}\,\frac{Sw^2}{m}\right) - q\left(\rho' Sb^2 + \frac{\rho'' r}{1-\rho''^2 s^2}\,\frac{Sw^2}{m}\right)\right]$$

$$\text{Cov}(\bar{x}_w, \bar{x}''_2) = \frac{\alpha}{n\triangle}\left[\alpha\left(Sb^2 + \frac{1-\rho''^2 s}{1-\rho''^2 s^2}\,\frac{Sw^2}{m}\right) - q\left(\rho' Sb^2 + \frac{\rho'' r}{1-\rho''^2 s^2}\,\frac{Sw^2}{m}\right)(\rho' Sb^2)\right]$$

$$\text{Cov}(\bar{x}_w, \bar{x}'_2) = \frac{\alpha}{n\triangle}\left[\alpha\left(Sb^2 + \frac{1-\rho''^2 s}{1-\rho''^2 s^2}\,\frac{Sw^2}{m}\right) - q\left(\rho' Sb^2 + \frac{\rho'' r}{1-\rho''^2 s^2}\,\frac{Sw^2}{m}\right)(\rho' Sb^2 + \rho''(Sw^2/m))\right].$$

This indicates that this estimator $\bar{x}_w$ cannot be more efficient than the estimator E_2 and thus the conclusion made in the paper referred to appears to be not correct. The Table 2 in the paper [Abraham *et.al.* (1969)] is corrected and is given here as Table 1. However, from comparison between the Table 1 in this paper and Table 1 in the paper referred to, it is observed that the superiority of estimator E_2 over $\bar{x}_w$ is not considerable.

It is assumed that the population parameters ρ', ρ'', Sb^2 and Sw^2 are known. In case they are not known, these may be calculated from the sample observations and to that extent the efficiency of the estimate is affected. The estimates of these population parameters are given below

$$\text{Est}(Sb^2_t) = sb^2_t - (sw^2_t)/m)$$

$$\text{Est}(Sw^2_t) = sw^2_t$$

$$\text{Est}(\rho'_{tt'}\, Sb_t\, Sb_{t'}) = r'_{tt'}\, sb_t\, sb_{t'} - (r''_{tt'}\, sw_t\, sw_{t'})/m$$

and

$$\text{Est}(\rho''_{tt'}\, Sw_t\, Sw_{t'}) = r''_{tt'}\, sw_t\, sw_{t'}$$

where

$$sb^2_t = (1/(np-1)) \sum_{k=1}^{np} (\bar{x}'_{tk} - \bar{x}'_t)^2$$

$$sw^2_t = (1/np(mr-1)) \sum_{k=1}^{np} \sum_{l=1}^{mr} (x_{tkl} - \bar{X}'_{tk})^2$$

$$r'_{tt'}\, sb_t\, sb_{t'} = (1/(np-1)) \sum_{k=1}^{np} (\bar{x}'_{tk} - \bar{x}'_t)\,(\bar{x}'_{t'k} - \bar{X}'_{t'})$$

TABLE 1

Efficiency of the estimator $\bar{x}_w$ over that of the sample mean per *ssu* on the second occasion, based on all *nm* units for $m=4$ and for different values of ρ', ρ'', p, r and φ.

		$\varphi=0·1$									$\varphi=1·0$									$\varphi=10·0$								
ρ'	ρ''	0·5			0·7			0·9			0·5			0·7			0·9			0·5			0·7			0·9		
	p r	0·25	0·50	0·75	0·25	0·50	0·75	0·25	0·50	0·75	0·25	0·50	0·75	0·25	0·50	0·75	0·25	0·50	0·72	0·25	0·50	0·75	0·25	0·50	0·75	0·25	0·50	0·75
	0·25	1·03	1·03	1·03	1·03	1·04	1·04	1·04	1·04	1·04	1·03	1·03	1·04	1·04	1·04	1·05	1·05	1·06	1·06	1·02	1·03	1·04	1·04	1·07	1·09	1·12	1·15	1·17
0·4	0·50	1·04	1·04	1·04	1·04	1·05	1·05	1·05	1·05	1·05	1·04	1·05	1·05	1·05	1·06	1·06	1·08	1·08	1·08	1·03	1·05	1·06	1·08	1·11	1·12	1·20	1·23	1·22
	0·75	1·03	1·03	1·03	1·03	1·03	1·03	1·04	1·04	1·04	1·03	1·04	1·04	1·05	1·05	1·05	1·08	1·08	1·07	1·04	1·06	1·06	1·09	1·12	1·11	1·24	1·25	1·20
	0·25	1·09	1·09	1·09	1·09	1·09	1·09	1·09	1·10	1·10	1·06	1·07	1·08	1·08	1·09	1·10	1·10	1·11	1·12	1·03	1·04	1·06	1·05	1·09	1·11	1·14	1·19	1·21
0·6	0·50	1·11	1·11	1·11	1·11	1·11	1·11	1·11	1·11	1·11	1·08	1·09	1·10	1·10	1·11	1·12	1·14	1·15	1·15	1·04	1·06	1·08	1·09	1·13	1·14	1·22	1·26	1·26
	0·75	1·07	1·07	1·07	1·07	1·08	1·08	1·08	1·08	1·08	1·06	1·07	1·07	1·08	1·09	1·09	1·12	1·12	1·11	1·05	1·07	1·07	1·10	1·13	1·13	1·25	1·27	1·22
	0·25	1·21	1·22	1·22	1·22	1·22	1·23	1·23	1·23	1·23	1·13	1·15	1·16	1·15	1·18	1·20	1·20	1·23	1·24	1·03	1·05	1·07	1·07	1·11	1·14	1·17	1·23	1·27
0·8	0·50	1·22	1·23	1·23	1·23	1·23	1·23	1·23	1·24	1·24	1·15	1·17	1·18	1·18	1·20	1·21	1·23	1·25	1·26	1·05	1·08	1·09	1·10	1·15	1·17	1·25	1·31	1·31
	0·75	1·14	1·14	1·14	1·14	1·14	1·14	1·15	1·15	1·15	1·11	1·12	1·12	1·13	1·14	1·14	1·17	1·18	1·17	1·05	1·08	1·08	1·11	1·15	1·15	1·27	1·29	1·25

and

$$r''_{tt'}\, sw_t\, sw_{t'} = \frac{1}{np(mr-1)} \sum_{k=1}^{np} \sum_{l=1}^{mr} (x_{tkl} - \bar{x}'_{tk})(x_{t'kl} - \bar{x}'_{t'k})$$

where x_{tkl} is the observation on character X of the lth *ssu* in kth *psu* on the tth occasion. Also

$$\bar{x}'_{tk} = (1/mr)) \sum_{l=1}^{mr} x_{tkl} \quad \text{and} \quad \bar{x}'_t = (1/(np)) \sum_{k=1}^{np} \bar{x}'_{tk}.$$

Summary

An estimation procedure has been discussed in two-stage successive sampling on the h occasions where primaries as well as secondaries are partially retained.

Acknowledgement

The authors are grateful to Dr. D. Singh, Director, Institute of Agricultural Research statistics for his constant encouragement and keen interest in the preparation of the paper. They are also thankful to the refree for his suggestions for the improvement of the paper.

References

1. Abraham, T.P., Khosla, R.K. and Kathuria, O.P (1969) : "Some investigations on the use of successive sampling in pest and diseases." Jour. Ind. Soc. of Agri. Stat., Vol 21, No 2.
2. Eckler, A.R. (1955) : "Rotation Sampling" Ann. of Math Stat., Vol. 26.
3. Jessen, R.J. (1942) : "Statistical investigations of sample survey for obtaining farm facts." Iowa Agri. Exp. Res. Bull. 304.
4. Patterson, H.D. (1950) : "Sampling on successive occasions with partial replacement of units." Jour. Royal Stat. Soc. Series B-12.
5. Singh, D. (1968) : "Estimation in successive sampling using a multistage design." Jour. Amer. Stat. Assoc. 63.
6. Singh, D. and Kathuria, O.P. (1969) : "On two-stage successive sampling." Aust. Jour. of Stat., Vol. 11, No 2.
7. Singh, S. (1970) : "Some studies in two-stage successive sampling." Thesis submitted for award of I.C.A.R.. Diploma—unpublished.
8. Tikkiwal, B.D. (1951) : "Theory of successive sampling." Thesis submitted for award of I.C.A.R. Diploma—unpublished.
9. _____ (1955) : "Multiphase sampling on successive occasions" Thesis submitted to the Faculty of North Carolina State College, U.S.A. in partial fulfilment of the requirements for the degree of Doctor of Philosophy.
10. _____ (1956) : "A further contribution to the theory of univariate sampling on successive occasions." Jour. of Ind. Soc. of Agri. Stat., vol. 8
11. _____ (1964) : "A note on two stage sampling on successive occasions." Sankhya, Series 4, Vol. 28, part I.
12. _____ (1965) : "The theory of two-stage sampling on successive occasions." Jour. of Ind. Stat. Assoc.
13. Yates, F. (1949) : "Sampling Methods for Census and Surveys." Charles Griffin and Company Ltd., London.