AN APPROACH FOR ESTIMATING CHANGES IN DYNAMIC POPULATIONS

By

J. P. Jain and M. Rajagopalan

Indian Agricultural Statistics Research Institute, New Delhi

(Received: October, 1976)

1. Introduction

When a survey is to be repeated over time recourse is taken to successive sampling, where the sampling units are partially matched from occasion to occasion. In addition, if the survey is to bring out the changes that have taken place due to the introduction of new technology, it is necessary to study the changes simultaneously in a 'control area' where the technology has not been introduced. The comparative picture of changes in the development area and the control area provides a measure of impact of the new technology.

The theory of successive sampling technique as developed by several workers such as Jessen [2], Patterson [3], Eckler [1], Rao and Graham [5], Tikkiwal [7, 8, 9, 10, 11, 12], Singh and Kathuria [6] and Kathuria and Singh [4] assumes that the size and structure of the population remains the same over time. Very often, the structure of the population changes to such an extent that the available methodology is not strictly applicable. For instance, in the study of impact of milk supply schemes on rural economy in milk collection areas the population of villages categorised as supplying milk to the organised agencies and those non-supplying may undergo a change in their status and composition partially or wholly over time. In other words, the number of villages in each category may change their status and the number of households in a village supplying milk may also vary from occasion to occasion. Due to these changes, the sample selected during the first occasion necessarily undergoes a change in its denomination over the subsequent occasions. To deal with this problem a suitable methodology needs to be developed and is discussed in this paper.

2. Sampling Design

A typical design adopted in such surveys is one of strartified two-stage random sampling. The National Extension Service Blocks, or Community Development Blocks or groups of taluks in each of the supplying and non-supplying areas, constitute the strata, with clusters of villages within each stratum as the first-stage units and households within a cluster as the second-stage units.

The design for the repeat survey would be the same as that for the survey on the first occasion. The pattern of selection of units in the still milk supplying area is to retain some of the psu's (clusters of villages) with their samples of ssu's (households) canvassed on the first occasion and to select some psu's afresh, whereas in the non-supplying area all the psu's canvassed on the first occasion, which are still non-supplying at the time of repeat enquiry, are retained with their sample psu's. The reason for not selecting a sample of psu's afresh from the non-supplying area is that we are not directly interested in obtaining the estimate on the second occasion for this area. The diagramatic representation of the villages so selected as psu's is given in the next section.

3. Approach to the Problem

In a situation described in the previous sections, the population could be conceived of as being split up into four sub-populations consisting of clusters of villages which were (*i*) supplying on both the occasions (designated as S/S) ; (*ii*) supplying on the first and non-supplying on the second occasion (S/N) ; (*iii*) non-supplying on the first and supplying on the second occasion (N/S) ; and (*iv*) non-supplying on both the occasion (N/N). Accordingly, the distribution of selected clusters of villages over two occasions in the four sub-populations can be shown as in the diagram given below :

Diagramatic representation of samples of clusters

	Sub-population		*Sampling pattern*		
1.	S/S	B	****	***	
		R		***	**
2.	S/N	B		**	
		R		**	
3.	N/S	B		***	
		R		***	**
4.	N/N	B		**	
		R		**	

B = Benchmark survey (first occasion),

R = Repeat survey (second occasion)

By restructuring the population as above, it would be possible to obtain an improved estimator of the character under study for the repeat survey for the sub-population 1 and simple estimates for others.

On the first occasion, the first two sub-populations namely, S/S and S/N together constitute the supplying area and the last two sub-populations namely, N/S and N/N, the non-supplying area as conceived at the first occasion. As such, the respective two sub-populations could be treated as one population for building up the estimates of the characters for the first occasion. As regards the estimates on the second occasion, the estimates of sub-populations 1 (S/S) and 3 (N/S) would be combined to obtain the estimate of the total of a character for the supplying area. Th ecorresponding estimate for the non-supplying area would be obtained likewise from sub-populations 2 (S/N) and 4 (N/N).

Since the foregoing estimates pertain to the overlapping populations on the two occasions, they will not provide valid estimate of change in the parameter between the two occasions. An appropriate estimate of change would be based on that part of the population which remained of the same denomination on both the occasions. In other words, an estimate of change would be arrived at by considering the first (S/S) and fourth (N/N) sub-populations only.

4. Notation

Let, for any given stratum

M_i : total nunber of clusters of villages in the i-th sub-population ($i=1, 2, 3, 4$) ;

m_{ij} : number of clusters sampled out of M_i on the j-th occasion ($j=1, 2$) ;

m'_i : number of selected matched ctusters in the i-th sub-population ;

m''_{ij} : number of selected unmatched clusters in the i-th sub-population on the j-th occasion ;

N_{ijk} : total number of households in the i-th sub-population on j-th occasion in the k-th selected cluster ;

N'_{ijk} : the corresponding total number of households in the k-th selected matched cluster ;

N''_{ijk} : the corresponding total number of households in the k-th selected unmatched cluster ;

n_{ijk} : number of households selected out of N_{ijk} from the i-th sub-population on j-th occasion in the k-th selected cluster ;

N'_{i*k} : number of common units between N'_{i1k} and N'_{i2k} ;

n_{i*k} : number of common units between n_{i1k} and n_{i2k} ;

x_{ijkl} : the value of a character recorded in the l-th selected household of the k-th cluster on the j-th occasion in the i-th sub-population.

Further, for any character x, we define the estimate of total for the i-th sub-population on the j-th occasion as

$$\hat{X}_{ij.}=\frac{M_i}{m_{ij}}\sum_{k=1}^{m_{ij}}\hat{X}_{ijk}=M_i\hat{\bar{X}}_{ij.}$$

$$\text{where, } \hat{X}_{ijk.}=\frac{N_{ijk}}{n_{ijk}}\sum_{l=1}^{n_{ijk}}x_{ijkl}=N_{ijk}\hat{\bar{X}}_{ijk.}$$

The corresponding estimates of total based on matched units is

$$\hat{X}'_{ij.}=\frac{M_i}{m'_i}\sum_{k=1}^{m'_i}\hat{X}'_{ijk.}=M_i\hat{\bar{X}}'_{ij.}$$

$$\text{where, } \hat{X}'_{ijk.}=\frac{N'_{ijk}}{n_{i*k}}\sum_{l=1}^{n_{i*k}}x_{ijkl}=N'_{ijk}\hat{\bar{X}}'_{ijk.}$$

and that based on unmatched units is

$$\hat{X}''_{ij.}=\frac{M_i}{m''_{ij}}\sum_{k=1}^{m''_{ij}}\hat{X}''_{ijk}=M_i\hat{\bar{X}}''_{ij.}$$

$$\text{where, } \hat{X}''_{ijk.}=\frac{N''_{ijk}}{n_{ijk}}\sum_{l=1}^{n_{ijk}}x_{ijkl}=N''_{ijk}\hat{\bar{X}}''_{ijk.}$$

5. Estimates of Total of a Character and Estimatfs of Variances

5·1. The total of a character for each of the sub-populations excepting for the first sub-population on the second occasion ($i\pm1$ and $j\pm2$) is obtained as a simple estimate and is as follows :

$$\hat{X}_{ij.}=\frac{M_i}{m_{ij}}\sum_{k=1}^{m_{ij}}\hat{X}_{ijk.} \qquad \text{...(5·1·1)}$$

The estimate of variance of $\hat{X}_{ij.}$ for any $i \neq 1$ and $j \neq 2$ is

$$\hat{V}(\hat{X}_{ij.})=M_i^2\left[\left(\frac{1}{m_{ij}}-\frac{1}{M_i}\right)s_{ij}^2+\frac{1}{m_{ij}M_i}\sum_{k=1}^{m_{ij}} N_{ijk}^2\left(\frac{1}{n_{ijk}}-\frac{1}{N_{ijk}}\right)s_{ijk}^2\right] \quad ...(5{\cdot}1{\cdot}2)$$

where, $s_{ij}^2=\frac{1}{(m_{ij}-1)}\sum_{k=1}^{m_{ij}}\left(\hat{X}_{ijk.}-\hat{\bar{X}}_{ij.}\right)^2$

and $s_{ijk}^2=\frac{1}{(n_{ijk}-1)}\sum_{l=1}^{n_{ijk}}\left(x_{ijkl}-\hat{\bar{X}}_{ijk.}\right)^2$.

5·2. The total of a character for the sub-population 1 on the second occasion is obtained as an improved linear unbiased estimator as

$$\hat{X}_{12.}=u(\hat{X}'_{11.}-\hat{X}''_{11.})+v(\hat{X}'_{12.}-\hat{X}''_{12.})+\hat{X}''_{12.} \quad ...(5{\cdot}2{\cdot}1)$$

where u and v are some constants obtained in a manner such that $V(X_{12.})$ is minimum and their values are given by

$$u=-\frac{\alpha'_{12}\,\beta'_{112}}{\gamma},\quad v=\frac{\alpha''_{12}\,\alpha_{11}}{\gamma}$$

$$\alpha_{1j}=\alpha'_{1j}+\alpha''_{1j} \text{ for } j=1,2$$

$$\gamma=\alpha_{11}\alpha_{12}-\beta'_{112}$$

$$\alpha'_{1j}=\hat{V}(\hat{X}'_{1j.}),\ \alpha''_{1j}=\hat{V}(\hat{X}''_{1j.}),\ \beta'_{112}=\hat{Cov}(\hat{X}'_{11.},\hat{X}'_{12.})$$

$$\hat{V}(\hat{X}'_{1j.})=M_1^2\left[\left(\frac{1}{m'_1}-\frac{1}{M_1}\right)s_{1j}^{2\prime}+\frac{1}{m'_1M_1}\sum_{k=1}^{m'_1} N_{1jk}^{\prime 2}\left(\frac{1}{n_{1jk}}-\frac{1}{N'_{1jk}}\right)s_{1jk}^{2\prime}\right]$$

$$\hat{V}(\hat{X}''_{1j.})=M_1^2\left[\left(\frac{1}{m''_{1j.}}-\frac{1}{M_1}\right)s_{1j}^{2\prime\prime}+\frac{1}{m''_{1j}M_1}\sum_{k=1}^{m''_{1j}} N_{1jk}^{\prime\prime 2}\left(\frac{1}{n_{1jk}}-\frac{1}{N''_{1jk}}\right)s_{1jk}^{2\prime\prime}\right]$$

$$\hat{C}ov(\hat{X}'_{11\cdot}, \hat{X}'_{12\cdot}) = M_1^2 \left[\left(\frac{1}{m'_1} - \frac{1}{M_1}\right) s'_{1b12} + \frac{1}{m'_1 M_1} \sum_{k=1}^{m'_1} N'^2_{1*k} \left(\frac{1}{n_{1*k}} - \frac{1}{N'_{1*k}}\right) s'_{1wk12}\right]$$

$$s'^2_{1j} = \frac{1}{(m'_1 - 1)} \sum_{k=1}^{m'_1} (\hat{X}'_{1jk\cdot} - \bar{\hat{X}}'_{1j\cdot})^2$$

$$s''^2_{1j} = \frac{1}{(m''_{1j} - 1)} \sum_{k=1}^{m''_{1j}} (\hat{X}''_{1jk\cdot} - \bar{\hat{X}}''_{1j\cdot})^2$$

$$s'^2_{1jk} = \frac{1}{(n_{1*k} - 1)} \sum_{l=1}^{n_{1*k}} (x_{1jkl} - \bar{\hat{X}}'_{1jk\cdot})^2$$

$$s''^2_{1jk} = \frac{1}{n_{1jk} - 1} \sum_{l=1}^{n_{1jk}} (x_{1jkl} - \bar{\hat{X}}''_{1jk\cdot})^2$$

$$s'_{1b12} = \frac{1}{(m'_1 - 1)} \sum_{k=1}^{m'_1} (\hat{X}'_{11k\cdot} - \bar{\hat{X}}'_{11\cdot})(\hat{X}'_{12k\cdot} - \bar{\hat{X}}'_{12\cdot}), \text{ and}$$

$$s'_{1wk12} = \frac{1}{(n_{1*k} - 1)} \sum_{l=1}^{n_{1*k}} (x_{11kl} - \bar{\hat{X}}'_{11k\cdot})(x_{12kl} - \bar{\hat{X}}'_{12k\cdot}).$$

The estimate of variance of $\hat{X}_{12\cdot}$ is approximately (assuming large sample) given by

$$\hat{V}(\hat{X}_{12\cdot}) = \alpha''_{12} \left[1 - \frac{\alpha''_{12}\alpha_{11}}{\gamma}\right] \quad \ldots(5\cdot2\cdot2)$$

where, α_{12}'', α_{11} and γ are as defined earlier.

6. Estimate of Change of Total of a Character

The index of change could be conceived of as the difference between the increase on the second occasion in production level in the supplying area over that in the non-supplying area. As mentioned earlier, the study on change will be based on that part of the population which remained common on both the occasions. Accordingly, the expression for the change would be

$$\hat{I}=(\hat{X}_{12\cdot}-\hat{X}_{11\cdot})-\frac{M_1}{M_4}(\hat{X}_{42\cdot}-\hat{X}_{41\cdot}) \qquad \ldots(6\cdot1)$$

The estimate of variance of $\hat{I}$ is given by

$$\hat{V}(\hat{I})=\hat{V}(\hat{X}_{12\cdot})+\hat{V}(\hat{X}_{11\cdot})+\left(\frac{M_1}{M_4}\right)^2\hat{V}(\hat{X}_{42\cdot})+\left(\frac{M_1}{M_4}\right)^2\hat{V}(\hat{X}_{41\cdot})$$

$$-2\hat{C}_{ov}(\hat{X}_{12\cdot},\ \hat{X}_{11\cdot})-2\frac{M_1^2}{M_4^2}\hat{C}_{ov}(\hat{X}_{42\cdot},\ \hat{X}_{41\cdot}) \quad \ldots(6\cdot2)$$

where,

$$\hat{C}_{ov}(\hat{X}_{12\cdot},\ \hat{X}_{11\cdot})=-u\alpha''_{11}$$

$$\hat{C}_{ov}(\hat{X}_{42\cdot},\ \hat{X}_{41\cdot})=M_4^2\left[\left(\frac{1}{m'_4}-\frac{1}{M_4}\right)s_{4b12}+\frac{1}{m'_4M_4}\sum_{k=1}^{m'_4}N'^2_{4*k}\left(\frac{1}{n_{4*k}}-\frac{1}{N'_{4*k}}\right)s_{4wk12}\right]$$

$$s_{4b12}=\frac{1}{(m'_4-1)}\sum_{k=1}^{m'_4}(\hat{X}_{41k\cdot}-\hat{\bar{X}}_{41\cdot})(\hat{X}_{42k\cdot}-\hat{\bar{X}}_{42\cdot})$$

$$s_{4wk12}=\frac{1}{(n_{4*k}-1)}\sum_{l=1}^{n_{4*k}}(x_{41kl}-\hat{\bar{X}}_{41k\cdot})(x_{42kl}-\hat{\bar{X}}_{42k\cdot}),$$

and other quantities are as defined earlier.

7. Illustration

With the object of studying the impact of milk supply schemes on rural economy in milk collection areas of Delhi Milk Scheme, the Institute of Agricultural Research Statistics carried out benchmark and repeat surveys in the rural tracts of Meerut, Bullandshahr and Gurgaon districts during 1966-67 and 1972-73 respectively. The sampling plan adopted for the survey was one of stratified two-stage random sampling with districts Meerut and Bullandshahr constituting the first stratum and district Gurgaon as the second stratum. Clusters of three villages each within a radius of 5 km constituted the first-stage units and households within clusters as the second-stage units. Further, the enquiry was carried out simultaneously in villages which supplied milk to Delhi Milk Scheme and in the non-supplying villages. During benchmark survey, 72 clusters, 51 from the supplying area and 21 from the non-supplying area were covered. The design adopted on the second occasion involved partial replacement of the sampling units selected on the first occasion. From the distribution of clusters of villages it was noted that some of the villages which were classified as non-supplying at the time of benchmark survey became supplying at the time of repeat survey and *vice-versa*. In the population, of the total number of clusters which were supplying milk at the time of benchmark survey 249 continued to supply milk during the repeat survey and 172 ceased to supply milk. Likewise, out of the total number of 1,238 non-supplying clusters, 283 supplied milk at the time of repeat survey. Further, the number of milk producer households of the commercial type in each cluster also changed. Thus, there occurred a change both in size and structure of the population. To deal with this situation, the approach outlined in the earlier sections was adopted. Table 1 gives the

TABLE 1

Estimates of buffalo milk production (tonnes) per day in commercial milk producer households on the two occasions

Sub-population	*Occasion*	*Estimate*	*%S. E.*
S/S	B R	119.412 124.737	6.99 5.24
S/N	B R	65.898 58.549	15.99 16.09
N/S	B R	20.061 58.798	26.38 12.86
N/N	B R	183.362 279.265	16.40 16.09
Supplying area	B R	185.010 183.535	7.21 5.48
Non-supplying area	B R	203.423 338.114	15.01 16.09
	Impact.....................	(−) 20.577	88.84

B : Benchmark survey ; R : Repeat survey estimates in respect of total buffalo milk production per day in commercial milk producer households on the two occasions and as also the estimate of impact of Delhi Milk Scheme.

From the estimates of daily production of buffalo milk based on the entire population as it existed on the respective two occasions and as also from those based on a part of the population which was of the same denomination on both the occasions, it is seen that the production remained almost static from the first to second occasion in the supplying area but it increased in the non-supplying area by about 60 per cent. The results also show that the Delhi Milk Scheme does not seem to have provided an impact in increasing milk production in the commercial households of the rural areas covered under the scheme. Inspite of this, more than 25 per cent of the villages came within the fold of DMS since the commencement of the benchmark survey.

Summary

In repeated sampling enquiries application of method of successive sampling with partial replacement of units is advantageous. Its use is, however, restricted to situation where the size and structure of the populations remains the same from occasion to occasion. Since these conditions are no longer tenable for dynamic populations, there is a need for developing new procedures of estimation. A rigorous methodology for dealing with such populations seems quite involved. In this paper a simple heuristic approach is indicated for estimating the totals of a character on the two occasions as well as the change that has taken place over time. The formulae developed have been illustrated with the data collected from the rural areas covered under the Delhi Milk Scheme during 1966-67 and 1972-73.

Acknowledgement

The authors are grateful to Dr. Daroga Singh, Director, Institute of Agricultural Research Statistics, New Delhi for his keen interest in the investigation.

References

[1] Eckler, A. R. (1955) : Rotation Sampling. *Ann. Math. Stato* ; 26, 664-85.

[2] Jessen, R. J. (1942) : Statistical investigations of a sample survey for obtaining farm facts. *Iowa Agri. Exp. Stat. Res. Bull.* 304.

[3] Patterson, H. D. (1950).: Sampling on successive occasions with partial replacement of units. *Jour. Roy. Stat. Soc. B*, 12, 241-55.

[4] Kathuria, O. P. and Singh, D. (1971). : Comparisons of estimates in two-stage sampling on successive occasions. *Jour. Soc. Agri. Stat.* 23, 31-51.

[5] Rao, J. N. K. and Graham, J. E. (1964) : Rotation designs for sampling on repeated occasions. *Jour. Amer. Stat. Assoc.*, 59, 492-509.

[6] Singh, D. and Kathuria, O.P. (1969) : On two-stage successive sampling. *Austral. Jour. Stat.*, 11, 59-66.

[7] Tikkiwal, B. D. (1975) : Theory of successive sampling. Thesis submitted for award of I. C. A. R. Diploma (unpublished).

[8] ——— (1953) : Optimum allocation in successive sampling. *Jour. Ind. Soc. Agri. Stat.*, 5, 101-102.

[9] ——— (1956) : A further contribution to the theory of univariate sampling on successive occasions. *Jour. Ind. Soc. Agri. Stat.*, 8, 84-90.

[10] ——— (1964) : A note on two-stage sampling on successive occasions. *Sankhya*, Series A, 26, Part 1, 97-100.

[11] ——— (1965) : The theory of two-stage sampling on successive occasions. *Jour. Ind. Stat. Assoc.*, Vol. 3, No. 2 and 3, April-July, 125-135.

[12] ——— (1967) : Theory of multiphase sampling from a finite or an infinite population on successive occasions. *Review Int. Stat. Instt.*, Vol. 35.