A STUDY OF NON-RESPONSE IN SUCCESSIVE SAMPLING

By

D. Singh, R. Singh and Padam Singh

Institute of Agricultural Research Statistics

(Received in May, 1973; Accepted in January, 1974)

1. Introduction

1.1 For dynamic populations which are subject to change from time to time, it is necessary to repeat surveys at a fixed regular interval of time in order to know the changes being brought about in the value of the character. Jesson (1942) initiated the use of ancillary information available on previous occasion to improve the estimate for the current occasion. Since then a lot of work has been done in successive sampling by many research workers.

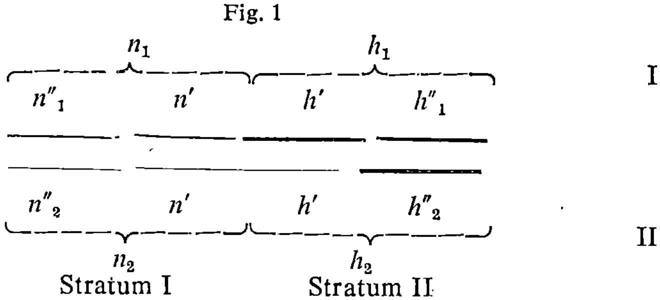
1.2 But the failure to obtain response from every member of the sample on any occasion will invalidate the procedures outlined so far since bias may be introduced in the estimators because of incomplete sample. In theory it can be eliminated by repeated call backs but in practice it is seldom possible because repeated calls on the one hand increase the cost of survey and on the other hand it delay its completion. Hansen and Harwitz (1946) were the first to deal with the problem of incomplete samples in mail surveys. However, in most instances mail surveys are not practicable. The skill and abilities of experienced enumerators are needed to obtain information sought in surveys. Bartholomew (1961) has suggested method by which almost unbiased estimates may be obtained after only two calls in interview surveys. The assumption being that the 2nd call sample will be random sample of all respondents not found at home at the first call, this being possible because in the course of first call the interviewer will be able to get information from neighbours or other members of the family and thus will be able to make some sort of appointment for the second call in such a way that the units missed in the first call have the same probability of being contacted in the second call.

In this paper our purpose is to extend this technique of Bartholomew for simple case of successive sampling on two occasions to obtain unbiased estimate in the presence of non-response from some units in the sample.

2. Sampling for two occasions

2.1 Consider a population of size N, where N is assumed to be very large. Let the sample size on two occasions be 'n' of which np units are common on both occasions. Now on first occasion let n_1 units respond on the first call and from the remaining $n-n_1$ (say m_1), let h_1 respond on the second call so that the sample size on the first occasion is n_1+h_1 which may be equal or less than n. On second occasion we are to have np units common to first occasion and nq units are to be selected afresh so that $p+q=1$. Now the np common units may be obtained proportionately from two components of the sample such that $n' \left(= \dfrac{n_1}{n_1+h_1} \cdot np \right)$ come from n_1 and h' $\left(= \dfrac{h_1}{n_1+h_1} \cdot np \right)$ come from h_1 units and we may further assume here that response is complete for these units. Now from the nq units to be selected afresh on the 2nd occasion let n''_2 respond on the first call and from the remaining $nq-n''_2$ (say m_2) let h''_2 units respond on the second call so that the total sample size on the second occasion is made up of four components viz., n', h', n''_2 and h''_2. Thus our sampling design will be as given in fig 1.

Fig. 1



Now the whole sampling design on the two occasions can be treated as consisting of two strata, the first stratum consisting of units responding on first call and second consisting of units responding on second call.

It may be stated here that the quantities n''_1, n''_2, h''_1 and h''_2 will generally be of random nature but for simplicity these have been treated as constant. Bartholomew (1961) has also treated them as constant while developing the sampling scheme.

2.2 *Estimate and Variance of Population Mean*

Let x_i and y_i denote the values of the character under study for the i-th unit on the first and the second occasion respectively.

Now for the units responding on the first call we define:

$\bar{x}_{n1}$ = Mean for the first occasion based on n_1 units responding on the first call.

$\bar{x}'_n$ = Mean for the first occasion based on n' units common to two occasions.

$\bar{y}_n'$ = Mean for the second occasion based on common units.

$\bar{x}_n{''}_1$ = Mean for the first occasion based on uncommon units.

$\bar{y}_n{''}_2$ = Mean for the second occasion based on uncommon units.

Similary, we can define corresponding estimates for the units responding on the second call just by replacing n by h.

Now we obtain an estimate of population mean on the second occasion using the auxiliary information available on the first occasion as suggested by Jesson (1942) for the first stratum as

$$\bar{y}_1 = \phi_1[\bar{y}_n' + \beta_1(\bar{x}_{n1} - \bar{x}_n')] + (1-\phi_1)\bar{y}_{n2}'' \qquad \ldots(1)$$

where

$$\phi_1 = \frac{n''_2}{n_2} \frac{\left(1 - \frac{n''_1}{n_1}\rho_1^2\right)}{\left(1 - \frac{n''_1 n''_2}{n_1 n_2}\rho_1^2\right)}$$

This is an unbiased estimate and its variance is given by

$$V(\bar{y}_1) = \phi_1 \frac{S_1^2}{n''_2}$$

where S_1^2 is the variance for first stratum assumed to be same for the two occasions. Similarly, another estimate may be obtained from the second stratum as

$$\bar{y}_2 = \phi_2[\bar{y}'_h + \beta_2(\bar{x}_{h_1} - \bar{x}_h')] + (1-\phi_2)\bar{y}_{h2}'' \qquad \ldots(2)$$

and

$$V(\bar{y}_2) = \phi_2 \frac{S_2^2}{h_2''}$$

where

$$\phi_2 = \frac{h''_2}{h_2} \frac{\left(1 - \frac{h''_1}{h_1}\rho^2_2\right)}{\left(1 - \frac{h''_1 h''_2}{h_1 h_2}\rho_2^2\right)}$$

and S_2^2 is the variance of second stratum assumed to be same for the two occasions.

Now we make the following assmuption :

(*i*) $\rho_1=\rho_2$ or $\beta_1=\beta_2$

(*ii*) $\phi_1=\phi_2$ which implies

$$\frac{n''_1}{n_1}=\frac{h''_1}{h_1} \quad \text{and} \quad \frac{n''_2}{n_2}=\frac{h''_2}{h_2}$$

We can consider a pooled estimate using $\bar{y}_1$ and $\bar{y}_2$ and obtain its variance expression in a simple and meaningful form.

Let

$$\bar{y}=\frac{n_2}{n}\,\bar{y}_1+\left(1-\frac{n_2}{n}\right)\overline{y_2} \qquad \ldots(3)$$

It is unbiased estimate of $\overline{Y}$ and its variance is given by

$$V(\bar{y})=EV(\bar{y}/n_1)+VE(\bar{y}/n_1) \qquad \ldots(4)$$

Now

$$VE(\bar{y}/n_1)=\frac{S^2}{n}\left(\frac{1-\rho^2 q}{1-\rho^2 q^2}\right) \qquad \ldots(5)$$

which is same as given by Jesson (1942), when there is complete response on both occasions,

and

$$EV(\bar{y}/n_1)$$

$$=\left(1-\frac{n_2}{n}\right)^2 S'^2_2\,\frac{\left\{\left(\frac{1}{h'}-\frac{1}{m_1}\right)-\left(\frac{1}{h'}-\frac{1}{h_1}\right)\rho^2\right\}\left(\frac{1}{h''_2}-\frac{1}{m_2}\right)}{\left(\frac{1}{h'}-\frac{1}{m_1}\right)-\left(\frac{1}{h'}-\frac{1}{h_1}\right)\rho^2+\left(\frac{1}{h''_2}-\frac{1}{m_2}\right)} \qquad \ldots(6)$$

This component is an addition in variance due to non-response, and it vanishes if—

(*i*) $n_2=n$ i.e. response is complete on the first call itself at the second occasion.

(*ii*) $h''_2=m_2$ i.e. response is complete on the second call at the second occasion.

An estimate of the additional term in variance due to non-response may be obtained by substituting an estimate of S_2^2, obtained from the units responding on the second call on the second occasion, and an estimate of ρ obtained from common units.

3. SAMPLING FOR TWO OCCASIONS
(*Finite Population*)

In section 2 we have assumed population size N to be very large. Even when population size N is not large, following Tikkiwal (1967), the estimator $\bar{y}$ given by (3) in section (2.2) remains valid. Variance of $\bar{y}$ then takes the form

$$V(\bar{y})=\{S^2/n\}.\ (1-\rho^2 q)/(1-\rho^2 q^2)-S^2/N$$
$$+\frac{\left(1-\frac{n_2}{n}\right)^2 S^2{}_2\left\{\left(\frac{1}{h'}-\frac{1}{m_1}\right)-\left(\frac{1}{h'}-\frac{1}{h_1}\right)\rho^2\right\}\left(\frac{1}{h''_2}-\frac{1}{m_2}\right)}{\left(\frac{1}{h'}-\frac{1}{m_1}\right)-\left(\frac{1}{h'}-\frac{1}{h_1}\right)\rho^2+\left(\frac{1}{h''_2}-\frac{1}{m_2}\right)} \quad ..\ (3.1)$$

Here we see that the addition in variance due to non-response is the same for finite as well as infinite population.

SUMMARY

In this paper an attempt has been made to handle the problem of non-response in successive sampling for two occasions. An unbiased estimate for the population mean on the second occasion and its variance expression have been obtained. It is found, as expected, that if the response is complete on the second occasion either in first call itself or in the first two calls, there is no increase in variance due to non-response. An estimate of the additional term in variance due to non-response has also been suggested.

ACKNOWLEDGEMENT

The authors are grateful to the editor for his suggestion how to remove the restriction of N being large for the results given in section 3.

REFERENCES

1. Bartholomew, D.J. (1961) : A method of allowing not at home bias in sample surveys. Applied Statistics, 10, 52-59.
2. Hansen, M.H. and Hurwitz, W.N. (1946) : The problem of non-response in sample surveys. J.A.S.A. 41, 517-529.
3. Jesson, R.J. (1942) : Statistical Investigation of a sample survey for obtaining farm facts. Iowa Agri. Exp. Stat. Res. Bull. 304.
4. Tikkiwal, B.D. (1954) : Theory of successive multiphase sampling (abstract) A.M.S. V. 25.
5. Tikkiwal, B.D. (1967) : Theory of Multiphase Sampling from a finite or an infinite population on successive occasions. Rev. of the International Stat. Instt. V. 35. 247-263.