MULTIPURPOSE SURVEYS ON SUCCESSIVE OCCASIONS

D. Singh and R. Singh

Institute of Agricultural Research Statistics, New Delhi-12

(*Received in October, 1971*)

1. Introduction

1.1. For dynamic populations which are subject to change from time to time such as area under improved seeds of crops, the extent of fertilizer's use etc., it is necessaJy to repeat surveys at a fixed regular interval of time in order to know the changes in the value of the character. Also while planning certain surveys, it is convenient and cheaper and sometimes it may be even necessary to include a number of related characters to be studied in a single enquiry. How many and which characters should be included in a survey, depends upon the scope and necessity of the survey. Once the characters to be included in a survey have been decided and the need for repeating such enquiries on successive occassions has been fully felt, the experimenters' interest would be to utilize the entire information to obtain most precise estimates of differet characters on different occasions.

1.2. The major problem that the experimenter faces in such situations is the lack of knowledge of the pattern of correlation between the values of a character on any two occasions. Patterson (1950) developed the method under the pattern that the correlation ρ, between units on successive occasions for a single character remains constant and correlation between units two occasions apart is ρ^2, that between the units three occasions apart is ρ^3, etc. The pattern of correlation assumed by Tikkiwal (1953) is slightly more general. While the correlation between units taken on different occasions has been allowed to vary, that between units on more than two occasions apart has been taken to be equal to the product of correlations between units on all pairs of consecutive occasions formed by these. Again for a multi-character study, Tikkiwal (1955) assumed the correlation between *i*-th character on *l*-th occasion and *j*-th character on *m*-th occasions as

$$\rho_{ij}^{(lm)} = \rho_{ij}^{l} \cdot \rho_{j}^{lm} ,$$

where

ρ_{ij}^{l} = correlation between i-th and j-th character on l-th occasion,

and ρ_{j}^{lm} = correlation between the values of j-th character on l-th and m-th occasion, and is the product of the series of correlations between the units of j-th character on two consecutive occasions from l to m.

1.3. In many situations, it is quite probable that the correlation between different characters on different occasions may not follow any specified pattern. It may be, then, desirable to use most general correlation pattern, say, $\rho_{h(k)r(s)}$', where $\rho_{h(k)r(s)}$ denotes the correlation coefficient between k-th character on h-th occasion and s-th character on r-th occasion, where

$$h, r = 1, 2, \ldots, t,$$

t being the number of times survey is repeated and

$$k, s = 1, 2, \ldots, c,$$

c being the number of characters observed on each occasion.

2. Sampling for two characters on two occasions

2.1. For simplicity we may consider the case of two characters on two successive occasions. Let N and n be the sizes of population and sample respectively on each occasion. On the second occasion, the sample consists of np units common to both occasions and nq units selected from $(N-n)$ units not observed on first occasion, where $p+q=1$. Thus, on each occasion we have two components in the sample and their means provide unbiased estimates of corresponding population means.

2.2. Let $x_{h(k)i}$ denote the value of i-th unit for k-th character on h-th occasion, then population mean of the k-th character on h-th occasion may be written as

$$\bar{X}_{h(k)} = \frac{1}{N} \sum_{1}^{N} x_{h(k)i} \qquad h, k = 1, 2.$$

Now an unbiased estimate for the population mean for 2nd character on 2nd occasion may be obtained as

$$\bar{x}_{2(2)} = I_1(\bar{x}'_{1(1)} - \bar{x}''_{1(1)}) + I_2(\bar{x}'_{1(2)} - \bar{x}''_{1(2)}) + I_3(\bar{x}'_{2(1)} - \bar{x}''_{2(1)}) + I_4\bar{x}'_{2(2)} + I_5\bar{x}''_{2(2)} \qquad \ldots(2.1)$$

or simply $\bar{x}_{2(2)} = LX'$,

where

L is row vector $(l_1, l_2, \ldots, l_5)$ and to be so chosen that

$l_4 + l_5 = 1$ and $V(\bar{x}_{2(2)})$ is minimum,

X is row vector

$[\bar{x}'_{1(1)} - \bar{x}''_{1(1)}), (\bar{x}'_{1(2)} - \bar{x}''_{1(2)}), (\bar{x}'_{2(1)} - \bar{x}''_{2(1)}), \bar{x}'_{2(2)}, \bar{x}''_{2(2)}]$,

where

$\bar{x}'_{h(k)}$ is the estimate of population mean for k-th character on h-th occasion based on units common to all occasions,

$\bar{x}''_{h(k)}$ is the corresponding estimate based on fresh units selected on h-th occasion only.

$(h, k = 1, 2)$

The variance of $\bar{x}_{2(2)}$ is given by

$$(l_1, l_2, l_3, l_4, l_5) \begin{bmatrix} V_{11} \cdots\cdots\cdots V_{15} \\ V_{21} \cdots\cdots\cdots V_{25} \\ \cdots\cdots\cdots\cdots\cdots \\ \cdots\cdots\cdots\cdots\cdots \\ V_{51} \cdots\cdots\cdots V_{55} \end{bmatrix} \begin{bmatrix} l_1 \\ l_2 \\ l_3 \\ l_4 \\ l_5 \end{bmatrix}$$

where $V_{ij} (= V_{ji})$ is defined as the covariance between terms in (2·1) having coefficients l_i and l_j.

The above expression for $V(\bar{x}_{2(2)})$ may be written as LAL' where A is the variance and covariance matrix (V_{ij}).

Now to minimize $V(\bar{x}_{2(2)})$, consider the function

$$F = LAL' - 2\lambda(LE' - 1),$$

where E is the row vector (0, 0, 0, 1, 1) and λ is undermined constant multiplier.

The minimum value of $V(\bar{x}_{2(2)})$ may be obtained by solving the equation

$$\frac{\partial F}{\partial L} = 0$$

or $\quad 2AL'-2\lambda E'=0$

Now

Let $\quad P=\frac{L}{\lambda}$ where P is vector defined by $(p_1, p_2, \ldots, p_5)$,

$$p_i=\frac{I_i}{\lambda}.$$

Then

$$AP'=E'$$

or $\quad P'=A^{-1}E'$

Since $\quad I_4+I_5=1,$

hence $\quad \lambda=\frac{1}{p_4+p_5}$

and $\quad L'=\lambda A^{-1}E'.$

Thus the best linear unbiased estimate of $\bar{X}_{2(2)}$ is given by $\bar{x}_{2(2)}=LX'$ and minimum variance is given by

$$V(\bar{x}_{2(2)})=LAL'=\lambda^2 EA^{-1}E' \qquad \ldots(2\cdot2)$$

Now, assuming simple random sampling from a large population, we may write

$$V(\bar{x}'_{h(k)})=\frac{S^2_{h(k)}}{np},$$

$$V(\bar{x}''_{h(k)})=\frac{S^2_{h(k)}}{nq},$$

$$\text{Cov}\,(\bar{x}'_{h(k)}, \bar{x}'_{r(s)})=\frac{S_{h(k)r(s)}}{np},$$

and

$$\text{Cov}\,(\bar{x}'_{h(k)}, \bar{x}''_{h(k)})=0=\text{Cov}\,(\bar{x}'_{h(k)}, \bar{x}''_{r(s)})$$ (since samples np and nq are independent).

where

$$S^2_{h(k)}=\frac{1}{N-1}\sum_{I}(x_{h(k)i}-\bar{X}_{h(k)})^2,$$

$$S_{h(k)r(s)}=\frac{1}{N-1}\sum_{i}(x_{h(k)i}-\bar{X}_{h(k)})(x_{r(s)i}-\bar{X}_{r(s)}).$$

Hence, the variance covariance matrix (A) can be written as

$$\begin{bmatrix} \frac{S^2_{1(1)}}{npq} & \frac{S_{1(1)1(2)}}{npq} & \frac{S_{1(1)2(1)}}{np} & \frac{S_{1(1)2(2)}}{np} & 0 \\ & \frac{S^2_{1(2)}}{npq} & \frac{S_{1(2)2(1)}}{np} & \frac{S_{1(2)2(2)}}{np} & 0 \\ & & \frac{S^2_{2(1)}}{npq} & \frac{S_{2(1)2(2)}}{np} & -\frac{S_{2(1)2(2)}}{nq} \\ & & & \frac{S^2_{2(2)}}{np} & 0 \\ & & & & \frac{S^2_{2(2)}}{np} \end{bmatrix}$$

Substituting the value of A in (2·2) we can obtain $V(\bar{x}_{2(2)})$. Also

$$\text{Est } (S^2_{h(k)}) = s^2_{h(k)} = \left(\frac{1}{n-1}\right) \sum_i (x_{h(k)i} - \bar{x}_{h(k)})^2$$

and

$$\text{Est } (S_{h(k)r(s)}) = s_{h(k)r(s)} = \left(\frac{1}{np-1}\right) \left(\sum_i (x_{h(k)i} - \bar{x}'_{h(k)})(x_{r(s)i} - \bar{x}'_{r(s)}) \right)$$

Substituting these estimates in (A) in (2·2) we obtain an estimate of $V(\bar{x}_{2(2)})$.

3. Sampling for k (>2) characters on h (>2) occasions

3.1. When the study is to be continued for more than two occasions, the sample may be drawn in several alternatives ways. We may consider the following alternative designs of sampling.

(*i*) If n be the sample size on each occasion, np sampling units may be common to all the occasions, while fresh sample of nq sampling units be selected from the remaining $(N-n)$ units on each occasion.

(*ii*) Between any two consecutive occasions np units may be common by sub-sampling np unit from nq units studied on the previous occasion. Here obviously $nq \geqslant np$,

(*iii*) Each time np units are retained by sub-sampling from the whole sample, n.

3.2. We shall be giving here the results for the sampling pattern (*i*) only. But these can be easily extended to other sampling schemes also with only minor alterations.

Now the estimate of population mean for k-th character studied on h-th occasion may be written as

$$\bar{x}_{h(k)} = LX'$$

where

L is row matrix $(I_1, I_2 \ldots\ldots, I_{hk}, I_{hk+1})$ and to be so choosen that

$$I_{hk}+I_{hk+1}=1 \text{ and } V(\bar{x}_{h(k)}) \text{ is minimum,}$$

X is row matrix $[(\bar{x}'_{1(1)}-\bar{x}''_{1(1)}), (\bar{x}'_{1(2)}-\bar{x}''_{1(2)}), \ldots, (\bar{x}'_{h(k-1)}-\bar{x}''_{h(k-1)}),$
$\bar{x}'_{h(k)}, \bar{x}''_{h(k)}]$

Now we may put $V(\bar{x}_{h(k)})=LAL'$

It will be minimum when L is given by

$$L'=\lambda A^{-1}E',$$

where E is row vector $(0, 0 \ldots 1, 1)$ of order $hk+1$,

$$\lambda=\frac{1}{p_{hk}+p_{hk+1}},$$

P is row vector $(p_1, p_2 \cdots p_{hk}, p_{hk+1})$ of order $hk+1$, $p_i=\frac{I_i}{\lambda}$. The minimum variance is given by

$$LAL'=\lambda^2 EA^{-1}E'$$

where

A may be written as

$$\begin{bmatrix} \frac{S^2_{1(1)}}{npq} & \frac{S_{1(1)1(2)}}{nqp} \cdots\cdots & \frac{S_{1(1)1(k)}}{npq} \cdots\cdots & \frac{S_{1(1)h(k)}}{np} & 0 \\ & \frac{S^2_{1(2)}}{npq} \cdots\cdots & \frac{S_{1(2)1(k)}}{npq} \cdots\cdots & \frac{S_{1(2)h(k)}}{np} & 0 \\ & & \cdots\cdots\cdots\cdots & \cdots\cdots\cdots\cdots & \cdots\cdots \\ & & & \cdots\cdots\cdots\cdots & \cdots\cdots \\ & & & & \frac{S^2_{h(k)}}{nq} \end{bmatrix}$$

3.3. Now we explain how these results can be extended for the sampling scheme (*ii*) given in section 3·2.

The estimate $\bar{x}_{h(k)}$ may be written as

$$\bar{x}_{h(k)}=LX'$$

where

L is row vector $(I_1, I_2. \ldots\ldots, I_{2k(h-1)+1})$ and is so choosen that

$$I_{2k(h-1)+1}+I_{2k(h-1)}=1,$$

X is row vector $((\bar{x}''_{1(1)}-\bar{x}'''_{1(1)}), \ldots (\bar{x}''_{1(k)}-\bar{x}'''_{1(k)}), (\bar{x}'_{2(1)}-\bar{x}'''_{2(1)}),$
$(\bar{x}''_{2(1)}-\bar{x}'''_{2(1)}), \ldots, \bar{x}'_{h(k)}, \bar{x}'''_{h(k)})$ of order $2k(h-1)+1$

where

$\bar{x}'_{h(k)}=$Estimate of population mean for k-th character on h-th occasion based on units common to h-th and $(h-1)$-th occasion. $h=2, 3, \ldots, t$

$\bar{x}''_{h(k)}$ = Corresponding estimate of population mean based on units which are common to h-th and $(h+1)$-th occasion, $h=1, 2, \ldots, (t-1)$

and $\bar{x}'''_{h(k)}$ = Corresponding estimate of population mean based on units which are observed on h-th occasion only, $h=1, 2, \ldots, t$.

and variance of $\bar{x}_{h(k)}$ is given by

$$V(\bar{x}_{h(k)})=LAL'=\lambda^2 EA^{-1}E',$$

where A is the variance-covariance matrix of order $2k(h-1)+1$ as defined earlier and can be obtained easily on similar lines, noting that covariance between $\bar{x}'_{h(k)}$ and $\bar{x}''_{h-1(k)}$ is $\frac{S_{h(k)h-1(k)}}{np}$ and all other covariance terms are 0 for $h=2, 3, \ldots, t$.

3.4. Thus it is seen that the results obtained here may also be applicable to other sampling schemes with minor adjustments in the estimate and its variance.

4. Sampling for a Two Stage Design

4.1. In this section the results of the previous section have been extended to a two stage sampling design.

Let N and M be the number of psu's and ssu's within each psu respectively in the population and n, m be the corresponding numbers in the sample so that sample size remains same over all occasions. For simplicity we shall assume that each psu has got the same number of ssu's and also the number of ssu's in the sample within each sampled psu is the same.

4.2. Now in a multistage sampling design, when the survey is repeated for more than two occasions, the sample may be drawn in many different ways. But in the present investigation results have been obtained for a particular sampling scheme where a simple random sample of np psu's with all their, m sampled ssu's have been retained from 1st occasion for all the subsequent occasions and it is supplemented on each occasion by a fresh sub-sample of nq psu's and sampling m ssu's within each selected psu so that sample size on each occasion remains the same (nm). Also we assume that N and M are large so that $\frac{1}{N}$ and $\frac{1}{M}$ may be ignored.

4.3. Let $x_{h(k)ij}$ denote the value of j-th ssu in i-th psu for k-th character on h-th occasion.

Now we define for the population

$$\bar{X}_{h(k)}=\frac{1}{NM}\sum_i \sum_j x_{h(k)ij}$$

$$\bar{X}_{h(k)i}=\frac{1}{M}\sum_{j} x_{h(k)ij}$$

and for the sample

$$\bar{x}_{h(k)i}=\left(\frac{1}{m}\right)\sum_{j} x_{h(k)ij},$$

$$\bar{x}_{h(k)}=\left(\frac{1}{mn}\right)\sum_{i}\sum_{j} x_{h(k)ij}=\left(\frac{1}{n}\right)\sum_{i} \bar{x}_{h(k)i},$$

$$\bar{x}'_{h(k)}=\left(\frac{1}{npm}\right)\sum_{i}\sum_{j} x_{h(k)ij},$$

$$\bar{x}''_{h(k)}=\left(\frac{1}{nqm}\right)\sum_{i}\sum_{j} x_{h(k)ij},$$

$$\bar{S}^2_{h(k)1}=\left(\frac{1}{N}\right)\sum_{i} S^2_{h(k)i}$$

$$=\frac{1}{N(M-1)}\sum_{i}\sum_{j}(x_{h(k)ij}-\bar{X}_{h(k)i.})^2$$

$$S_{h(k)r(s)1}=\frac{1}{N(M-1)}\sum_{i}\sum_{j}(x_{h(k)ij}-\bar{X}_{h(k)i.})(x_{r(s)ij}-\bar{X}_{r(s)i.})$$

$$S^2_{h(k)2}=\left(\frac{1}{N-1}\right)\sum_{i}(\bar{X}_{h(k)i.}-\bar{X}_{h(k)})^2$$

$$S_{h(k)r(s)2}=\left(\frac{1}{N-1}\right)\sum_{i}(\bar{X}_{h(k)i.}-\bar{X}_{h(k)})(\bar{X}_{r(s)i.}-\bar{X}_{r(s)})$$

also

$$\bar{s}^2_{h(k)1}=\frac{1}{n(m-1)}\sum_{i}\sum_{j}(x_{h(k)ij}-\bar{x}_{h(k)i.})^2$$

and similarly $s_{h(k)r(s)1}$, $s^2_{h(k)2}$, $s_{h(k)r(s)2}$ can be defined.

Now we can write further

$$\text{Est}\ \bar{S}^2_{h(k)1}=\bar{s}^2_{h(k)1}$$

$$\text{Est}\ S_{h(k)r(s)1}=s_{h(k)r(s)1}$$

$$\text{Est}\ S^2_{h(k)2}=\left(s^2_{h(k)2}-\frac{\bar{s}^2_{h(k)1}}{m}\right)$$

$$\text{Est } S_{h(k)r(s)2} = \left(s_{h(k)r(s)2} - \frac{S_{h(k)r(s)1}}{m} \right)$$

and
$$\text{Est } V(\bar{x}'_{h(k)}) = \frac{1}{np} \left(s^2_{h(k)2} + \frac{\bar{s}^2_{h'k)1}}{m} \right) = \frac{A^2_{h(k)}}{np} \quad \text{(say)}$$

$$\text{Est } V(\bar{x}''_{h(k)}) = \frac{A^2_{h(k)}}{nq} \quad \text{and}$$

$$\text{Est Cov } (\bar{x}'_{h(k)}, \bar{x}'_{r(s)}) = \frac{1}{np} \left(s_{h(k)r(s)2} + \frac{s_{h(k)r(s)1}}{m} \right) = \frac{A_{h(k)r(s)}}{np}$$

Now an unbiased minimum variance estimator for the mean of k-th character on h-th occasion may be obtained easily on similar lines as earlier for a single stage sampling design.

4.4. Even if the number of characters varies from one occasion to another which may happen because some new characters need be included or some other characters dropped during the period of survey, the formulaes and procedure will remain the same with a few minor modifications.

5. Example

5.1. The results have been applied to the data obtained in Agronomic and Agro-economic surveys of the IADP district, Aligarh. Different estimates of the consumption of Nitrogeneous fertilizer per cultivator for the year 1966-67, using auxiliary variates, (*i*) total irrigated area, (*ii*) total area under all crops, for the four years (1963-64 to 1966-67) have been obtained. The results are presented below:

Estimate	*Mean per cultivator in kg*	*Estimate of Variance*	*Percentage efficiency*
$\overline{Y}_1$	87·5325	63·7075	100
$\overline{Y}_2$	93·6563	57·6304	110·54
$\overline{Y}_3$	89·7030	51·5105	123·67
$\overline{Y}_4$	78·0350	49·3715	129·03
$\overline{Y}_5$	81·9701	46·8815	135·89

where

$\overline{Y}_1$=simple mean of consumption of nitrogenous fertilizer per cultivator for the year 1966-67

$\overline{Y}_2$=mean of consumption of nitrogenous fertilizer per cultivator for the year 1966-67 utilizing information available on previous two years (1964-65 to 1965-66), for the character under study only.

$\overline{Y}_3$=mean of consumption of nitrogenous fertilizer per cultivator for the year 1966-67 utilizing information on previous two years (1964-65 to 1965-66) for all the three characters.

$\overline{Y}_4$=estimate utilizing available information for all the three characters on previous three occasions.

$\overline{Y}_5$=estimate of mean utilizing complete available information *i.e.* information on all the previous occasions for all characters and on current occasion for rest of the characters only.

5.2. Thus, it is observed, as expected, that by including additional auxiliary variables and by utilizing information available from the previous enquiry the efficiency of the estimate is improved. It is encouraging to note that use of additional information improves the efficiency of the estimates consistently and such improvement is not in significant.

SUMMARY

The Paper gives the estimates of population mean for each of k characters in multipurpose surveys conducted on h successive occasions. No restrictions on the correlation pattern have been imposed. As obvious, efficiency of the estimate is increased by taking all available information for all occasions.

REFERENCES

1. Patterson, H.D. (1950) : Sampling on successive occasions with partial replacement of units, JRSS series B-12.

2. Singh, D. (1968) : Estimates in successive sampling using multistage design. Jour. of American State. Asson.

3. Tikkiwal, B.D. (1953) : Theory of successive sampling. Jour. Ind. Soc. Agri. Stat.-5.

4. Tikkiwal, B.D. (1955) : Multiphase sampling on successive occasions, thesis submitted for ward of Ph. D. degree to North Crolina State College, U.S.A.

5. Singh, R. (1970) : Contribution to successive sampling. Thesis approved for Diploma in Agricultural and Animal Husbandry Statistics at IARS(ICAR), New Delhi.

6. Expert Committee on Assessment and Evaluation, Ministry of Food and Agrl. (1962-63). : Report on Intensive Agricultural District Programme, 1962-63.