CONTRIBUTION TO SUCCESSIVE SAMPLING

By
RANDHIR SINGH

DISSERTATION SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF DIPLOMA IN AGRICULTURAL AND ANIMAL HUSBANDRY STATISTICS OF THE INSTITUTE OF AGRICULTURAL RESEARCH STATISTICS
(I.C.A.R)
NEW DELHI-12
1970

ACKNOWLEDGEMENTS

I have great pleasure in expressing my deepest sense of gratitude to Dr. D. Singh Director, Institute of Agricultural Research Statistics (I.C.A.R.), New Delhi for his valuable guidance, constant encouragement and constructive criticism during the course of investigation and writing this thesis.

My thanks are due to Indian Council of Agricultural Research for providing financial assistance in the form of fellowship.

Randhir Singh

(RANDHIR SINGH)

C O N T E N T S

CHAPTER - I

INTRODUCTION

Usually the results of a sample survey are useful for the occasion when the survey is conducted. But for a dynamic population, i.e. which is subject to change from time to time such as area under improved seeds, the extent of fertilizer used, area under irrigation or number of unemployed persons etc., it is desirable to repeat survey at some regular intervals of time and to obtain improved current estimates, the use of ancillary information obtained on some previous occasion should be considered. Also to study the problem of estimating changes from one occasion to the other occasion, taking place in the population or to study the average of population over a period of time it is essential that survey should be repeated and also the survey should be able to provide reasonably reliable estimates of the value of character under study for the period in which survey is conducted.

The duration between two successive surveys is to depend upon the type of surveyed material, the information that is to be collected and the expenditure that has been sanctioned for the surveys.

The study of optima for successive sampling presents some new aspects especially applicable to such problems only. For example the questions are raised whether the sampling units should be identical or they should be quite different at successive occasions, if not what proportion of units should be identical i.e., common to both occasions and how to utilize the information from previous survey to improve the estimates of current figures? A solution of problems for optimum allocation of sample for each of two occasions was first proposed by Jessen (1942) and extended by Patterson.

Given the data from a series of samples there are three kinds of quantities for which we may be interested to get estimates.

1. The change in Y, the mean of character under study, from one occasion to the enext occasion.
2. The average value of Y over all occasion or for a period of time.
3. The average value of Y for the most recent occasion.

Suppose we are free to alter or retain the composition of sample and that total size of sample is to be same on all occasions. If we wish to maximize the precision the following statements can be made about replacement policy.

1. For estimating change it is best to retain same sample throughout all the occ-asions.
2. For estimating average over all occasions it is best to draw a new sample on each

occasion.

3. For obtaining estimates for the current occasion equal precision is obtained in both ways.

Now for selecting a fresh sample on each occasion so many difficulties relating to field operation are confrunted with i.e. the enumerator being new to the place of inquiry, may not get full cooperation of local population secondly the whole procedure of preparing the basic frame, tabulating and listing of sampling units shall have to be repeated each time and thus consuming more time and causing the cost of enumeration to rise high.

If the sampling units are not varying much, a resurvey of same units may fail to give any additional information of any particular interest. Moreover as Yates (1949) puts it "a repeated survey of same units may result in modification of these units relating to the rest of population, he asserts the point by giving one example that in a survey of agricultural practices, visits to farms may result in the farmers concerned, improving their practices through the advice from the investigator, an advice which when asked for can scarcely be refused." Thus retaining some of units from previous enquiry and supplementing them with a sample selected a fresh from the population seems to be an effective policy for adoption for field work.

In a survey involving a single stage random sample design, partial replacement of sampling un its does not present any serious problem. But for multistage designs, problem, is not so simple and various problems arise, for example, for two stage designs, what fraction of primary stage units should be selected a fresh and what fraction to be retained from first occasion to the second occasion, then what fraction of secondry stage units with in selected primary stage units should be partially replaced and partially retained, all these problems arise.

While a good amount of work has been done for single stage sampling design by Jassen, Patterson, Tikkiwal etc. not much work has been done for multistage design, D. Singh (1968) initiated the work for two stage design, he studied the replacement policy for primary stage units from first occasion to second occasion while the same second stage units were observed for the selected first stage units and he presented the optimum design for two occasions. Singh & Kathuria (1968) also studied two stage design for aloccasions and they studied the policy of replacement of secondary stage units while studying same primary stage units from one occasion to next occasion.

In this work, the bias in the estimate of population mean of a single character studied utilizing information on the same character on a previous occasion has been examined. Bias in the estimate of variance of this estimate has also been examined.

Next, when many characters are observed in a single enquiry and the survey is conducted continuously on different occasions, the efficiency of an estimate on a particular occasion can be increased by utilizing the ancillary information available on the correlated characters on the same occasion as well as information for all the characters on all previous occasion. The estimate of mean for a general kth character on hth occasion and its variance have been calculated.

The results obtained may be applicable to many a surveys. For the numerical illustrations we have considered the data obtained through Agronomic and Agro-economic survey of I.A.D.P. District Aligarh from years 1963-64 to 1966-67.

C H A P T E R -II

SAMPLING OF A SINGLE CHARACTER

Let N be the size of population under study which is sampled repeatedly on two occasions. For the purpose of our study we assume throughout that N is very large. Let n be the sample size on 1st occasion. Now on second occasion we retain a sample of np units from the sample on the 1st occasion and supplement it by nq selected independently from the (N-n) units. We shall denote by X's the population characteristic on the 1st occasion and Y's the same on the second occasion.

Thus on each occasion we have two components in the sample and their means provide unbiased estimate of corresponding population means.

We define these estimates as follows.

$\bar{x}_1' = \frac{1}{np} \sum_{i=1}^{np} x_i$ = Mean on the 1st occasion based on np units common to second occasion.

$\bar{x}_2'' = \frac{1}{nq} \sum_{i=1}^{nq} x_i$ = Mean on 1st occasion based on nq units which are not included on 2nd occasion.

and $\bar{y}_1', \bar{y}_2''$ are defined similarly.

$\bar{x}$ = Mean based on n units for 1st occasion.

Now a linear unbiased estimates of $\bar{Y}_2$ the population mean on second occasion is given by

$$\hat{\bar{Y}}_2 = a_1(\bar{x}_1' - \bar{x}_1'') + a_2 \bar{Y}_2' + (1-a_2)\bar{Y}_2''$$

and its variance is given by

$$V(\hat{\bar{Y}}_2) = \frac{a_1^2 S_x^2}{npq} + a_2^2 \frac{S_y^2}{np} + \frac{(1-a_2)^2 S_y^2}{nq} + \frac{2a_1 a_2 S_{xy}}{np}$$

where S_x^2 is variance of X, S_y^2 is variance of Y and S_{xy} is the covariance between X & Y.

Now for minimising the variance, differentiating partially with respect to a_1 and a_2 and equating to zero we get

$$\frac{\partial V(\bar{Y}_2)}{\partial (a_1)} = 0$$

i.e. $$\frac{2a_1 S_x^2}{npq} + 2 a_2 \frac{S_{xy}}{np} = 0$$

or $$a_1 = - \frac{S_{xy}}{S_x^2} q a_2$$

and $$\frac{\partial V(\bar{Y}_2)}{\partial (a_2)} = 0$$

i.e. $$2a_2 \frac{S_y^2}{np} - 2(1-a_2)\frac{S_y^2}{nq} + 2a_1 \frac{S_{xy}}{np} = 0$$

Values of a_1 and a_2 which will minimize the variance of $\bar{Y}_2$ are given by

$$a_1 = -\frac{s_{xy}}{s_x^2}\, a_2\, q$$

$$a_2 = \frac{\rho}{1-\rho^2 q^2} \quad \text{So that } a_1 = \frac{-\rho q \rho}{1-\rho^2 q^2}$$

Now if we assume $s_x^2 = s_y^2$ then the minimum variance of $\bar{Y}_2$ is given by

$$V(\hat{\bar{Y}}_2) = \frac{s_y^2\,(1-\rho^2\, q)}{n(1-\rho^2 q^2)}$$

The estimate $\bar{Y}_2$ is unbiased only if β (the regression coefficient of Y on x) is known but generally in practice this is estimated by only the common units in the sample. The estimate is still unbiased if X and Y follow a bivariate normal distribution, but this is however obvious that this estimate is always consistant.

Now the estimate of variance is given by

$$\hat{V}\,(\hat{\bar{Y}}_2) = \frac{s_y^2\,(1-r^2\, q)}{n\,(1-r^2 q^2)}$$ which is not unbiased.

<u>Expression for bias in the estimate of variance of mean.</u>

The bias in estimating above variance is given by

$$E\,\hat{V}\,(\hat{\bar{Y}}_2) - V(\hat{\bar{Y}}_2)$$

Now E $\hat{V}(\hat{\bar{Y}}_2)$ is given by

$$E\left\{\frac{s_y^2}{n}\,\frac{(1-r^2q)}{1-r^2q^2}\right\}$$

$$= E\left\{\frac{s_y^2}{n}\,(1-r^2q)\,(1-r^2q^2)^{-1}\right\}$$

$$= E\left\{\frac{s_y^2}{n}\,(1-r^2q)\,(1+r^2q^2)\right\}$$ Neglecting the terms of high order of qr

$$= E\,\frac{s_y^2}{n}\,(1-r^2pq)$$

$$= E\,\frac{s_y^2}{n} - E\left(r^2s_y^2\,\frac{pq}{n}\right)$$

$$= \frac{s_y^2}{n} - \frac{pq}{n}\,E\,(r^2s_y^2)$$

Now V $(\hat{\bar{Y}}_2)$ may also be written as

$$\frac{S_y^2}{n}\,\frac{1-\rho^2q}{1-\rho^2q^2} = \frac{S_y^2}{n}\,(1-\rho^2q)(1-\rho^2q^2)^{-1}$$

$$= \frac{S_y^2}{n}\,(1-\rho^2q)\,(1+\rho^2q^2)$$ neglecting higher powers of ρq

$$= \frac{S_y^2}{n}\,(1-\rho^2\,pq)$$

Hence, bias $= \frac{s_y^2}{n} - \frac{pq}{n}\,E\,(r^2s_y^2) - \frac{S_y^2}{n}$

$$+ \rho^2\,S_y^2\,\frac{pq}{n}$$

$$= \frac{pq}{n} \quad s_y^2 \rho^2 = E(s_y^2 r^2)$$

2.1. To find $E(r^2 s_y^2)$

$$E(s_y^2 r^2) = E\left\{ \frac{s_y^2 s_{xy}^2}{s_x^2 s_y^2} \right\} = E \frac{s_{xy}^2}{s_x^2}$$

When we assume that x's on two occasions follow a bivariate distribution then the joint distribution of s_x^2, s_{xy}, s_y^2 has the frequency function F given by

$$F = \frac{n^{n-1}}{4\pi \sqrt{n-2}} \quad \frac{(m20\ m02 - m11^2)^{\frac{n-4}{2}}}{M^{\frac{n-1}{2}}}$$ 4x " Cramer

$$e^{-\frac{n}{2M}(\mu 02\ m20 - 2\mu 11\ m11 + \mu 20 m02)}$$

in the domain $m20 > 0$, $m02 > 0$ $m11^2 < m20\ m02$

F = 0 outside the domain

Here $m20 = s_x^2$, $m02 = s_y^2$, $M11 = s_{xy}$

$\mu 20 = S_x^2$, $\mu 02 = S_y^2$, $\mu 11 = S_{xy}$

$$M = \begin{vmatrix} \mu 02 & \mu 11 \\ \mu 11 & \mu 20 \end{vmatrix}$$

Now we make the following transformations

$$u = \frac{m11^2}{m20}$$

$$v = m20$$

$$w = m02$$

$$\frac{1}{J} = \begin{vmatrix} \frac{\partial u}{\partial m11} & \frac{\partial u}{\partial m20} & \frac{\partial u}{\partial m02} \\ \frac{\partial v}{\partial m11} & \frac{\partial v}{\partial m20} & \frac{\partial v}{\partial m02} \\ \frac{\partial w}{\partial m11} & \frac{\partial w}{\partial m20} & \frac{\partial w}{\partial m02} \end{vmatrix} = 2\sqrt{\frac{u}{v}}$$

Then $J = \frac{1}{2}\sqrt{\frac{v}{u}}$

Now $F(u,v,w) = C\, u^{-\frac{1}{2}}\, v^{\frac{n-3}{2}}\, (w-u)^{\frac{n-4}{2}}\, e^{-\frac{n}{2M}\left(\mu 02\, v - 2\mu 11\sqrt{uv} + \mu 20 w\right)}\, du.dv.dw$

Put $w-u = z$

$dw = dz$

Now integrating over z, F may be written as

$$F(u,v) = C\, u^{-\frac{1}{2}}\, v^{\frac{n-3}{2}}\, e^{-\frac{n}{2M}\left(\mu 02\, v - 2\mu 11\sqrt{uv}\right)}\, du.dv$$

$$\times \int_0^\infty z^{\frac{n-4}{2}}\, e^{-\frac{n}{2M}\mu 20\,(u+z)}\, dz$$

or $F(u,v) = C\, u^{-\frac{1}{2}}\, v^{\frac{n-3}{2}}\, e^{-\frac{n}{2M}\left(\mu 02\, v - 2\mu 11\sqrt{uv}\right)} \times$

$$\times \quad \frac{\overline{n-2}}{2} \left(\frac{n}{2M} \mu 20 \right)^{-\frac{n-2}{2}} e^{-\frac{n}{2M} \mu 20 u}$$

Now integrating w.r.t. v we get

$$F(u) = C \sqrt{\frac{n-2}{2}} \left(\frac{n}{2M} \mu 20 \right)^{-\frac{(n-2)}{2}} u^{-\frac{1}{2}} e^{-\frac{n}{2M} \mu 20 u}$$

$$\times \; v^{\frac{n-3}{2}} e^{-\frac{n}{2M} (\mu 02 v - 2\mu 11 \sqrt{uv})} \; dv \, du$$

Now the terms under integration may be written as

$$\int_0^\alpha v^{\frac{n-3}{2}} e^{-\frac{n}{2M} \mu 02 v} \sum_{r=0}^{\alpha} \left(\frac{n}{M} \mu 11 \right)^r u^{\frac{r}{2}} v^{\frac{r}{2}} \, dv$$

$$= \frac{\left(\frac{n}{M} \mu 11\right)^r}{\overline{r}} u^{\frac{r}{2}} \sqrt{\frac{n+r-1}{2}} \left(\frac{n}{2M} \mu 02 \right)^{-\frac{(n+r-1)}{2}}$$

Now $E(u) = \int u \, F(u)$

$$= C' \int_0^\alpha u^{\frac{r+1}{2}} e^{-\frac{n}{2M} \mu 20 u} \, du$$

$$= C \frac{\sqrt{\frac{n-2}{2}}}{\overline{r}} \left(\frac{n}{2M} \mu 20 \right)^{-\frac{n-2}{2}} \left(\frac{n}{M} \mu 11 \right)^r \left(\frac{n}{2M} \mu 02 \right)^{-\frac{n+r-1}{2}}$$

$$\times \sqrt{\frac{n+r-1}{2}} \left(\frac{n}{2M} \mu 20 \right)^{-\frac{r+3}{2}} \sqrt{\frac{r+3}{2}}$$

$$= \frac{n^{n-1} \sqrt{\frac{n-2}{2}} \sqrt{\frac{n+r-1}{2}} \sqrt{\frac{r+3}{2}}}{2\pi \, \overline{n-2} \; M^{\frac{n-1}{2}}} \left\{ \left(\frac{n}{2M} \mu 20 \right) \right\}^{-\frac{n+r+1}{2}}$$

$$\times \left(\frac{n}{2N} \mu 02 \right)^{-\frac{n+r-1}{2}} \left(\frac{n}{N} \mu 11 \right)^{r}$$

$$= \frac{\Gamma\frac{n-2}{2}\,\Gamma\frac{n+r-1}{2}\,\Gamma\frac{r+3}{2}}{\Gamma n+2\;\;\Gamma r+1}\; 2^{n+r-3}\, \rho^{r}\, (1-\rho^{2})^{\frac{n+1}{2}} \mu 20$$

2.2 Now putting the value of $E(s_y^2\, r^2)$ we get

relative bias = $\dfrac{\text{Bias in estimating variance}}{\text{variance}}$

$$= \frac{pq}{1-\rho^{2}\, pq} \left[\rho^{2} - \frac{(1-\rho^{2})^{\frac{n+1}{2}}}{n\;\pi} \times \right.$$

$$\left. \frac{\Gamma\frac{n-2}{2}}{\Gamma n-2} \sum_{r=0}^{\infty} \rho^{r}\; \frac{2^{n+r-3}\,\Gamma\frac{n+r-1}{2}\,\Gamma\frac{r+3}{2}}{\Gamma r+1} \right]$$

2.3 Let B_n be the relative bias when sample size is n and it is B_{n+2} when sample size is $(n+2)$.

$$\text{Then}\quad B_n - B_{n+2} = \frac{pq}{1-\rho^{2}pq} \left(\rho^{2} - \frac{(1-\rho^{2})^{\frac{n+1}{2}}}{n} \right) \frac{\Gamma\frac{n-2}{2}}{\Gamma n-2}$$

$$\sum_{r=0}^{\infty} \frac{2^{n+r-3}\,\Gamma\frac{n+r-1}{2}\,\Gamma\frac{r+1}{2}}{\Gamma r+1}\, \rho^{r} - \frac{pq}{1-\rho^{2}\, pq} \left[\rho^{2} \right.$$

$$\left. - \frac{(1-\rho^{2})^{\frac{n+3}{2}}}{\Gamma n+2} \; \frac{\Gamma\frac{n}{2}}{\Gamma n} \sum_{r=0}^{\infty} \rho^{r}\; \frac{\Gamma\frac{n+r+1}{2}}{\Gamma r+1}\; \Gamma\frac{r+1}{2}\; 2^{N+r-1} \right]$$

$$= \frac{pq}{1-\rho^2\, pq} \; (1-\rho^2)^{\frac{n+1}{2}} \sum_{r=0}^{\infty} \rho^r \; \frac{\frac{r+3}{2}}{\sqrt{r+1}} \; \frac{\frac{n+r-3}{2}}{\sqrt{n-2}} \; \sqrt{\frac{n-2}{2}}$$

$$\sqrt{\frac{n+r-1}{2}} \; x\left[4\;\frac{(1-\rho^2)}{n+2}\;\frac{n+r-1}{2}\;\frac{n-2}{2}\;\frac{1}{n-1}\;\frac{1}{n-2} - \frac{1}{n}\right]$$

$$= Qn \sum f(r,n)\left[(1-\rho^2)\;\frac{(n+r-1)}{(n+2)(n-1)} - \frac{1}{n}\right]$$

$$= Qn \sum f(r,n)\left[\frac{(1-\rho^2)}{n+2} + \frac{r(1-\rho^2)}{(n-1)(n-2)} - \frac{1}{n}\right]$$

$$= Qn \sum f(r,n)\left[r\;\frac{(1-\rho^2)}{(n-1)(n-2)} - \frac{1}{n} + \frac{1-\rho^2}{n+2}\right]$$

$$= Qn \sum_{r=0}^{\infty} f(r,n)\;(rA_n - C_n)$$

Where $A_n = \dfrac{1-\rho^2}{(n-1)(n-2)}$

$$C_n = \frac{1}{n} - \frac{1-\rho^2}{n+2}$$

$$= Qn\;A_n \sum_{r=0}^{\infty}\left[r\;f(r,n) - \frac{C_n}{A_n}\;f(r,n)\right]$$

So above expression is of the form

$$D_n \sum_{r=0}\left[r\,f(r,n) - B_n\,f(r,n)\right]$$

Where $B_n = \dfrac{C_n}{A_n}$ and $D_n = Q_n\,A_n$

As E_n is finite for given n and $f(r,n)$ is a function of r and n it is easy to see that

$$\sum_{r=0} r\, f(r,n) > E_n \sum_{r=0} f(r,n)$$

Hence $B_n - B_{n+2}$ is positive, showing that the relative bias of the estimate of variance is monotonically decreasing function of n.

CHAPTER - III

STUDY OF TWO CHARACTERS ON TWO OCCASIONS.

In practice we observe many characters in the same enquiry. If we are interested in estimating with maximum precision the population means of different characters on different occasions, the information available on all the correlated characters on all previous occasion can be utilized to increase the precision of the estimate.

We shall first discuss the study of two characters on two occasions.

We denote by $X_{h(k)i}$ the value of ith unit for kth character on hth occasion, h, k = 1,2, i=1, 2,.....n

Now as earlier we define

$$\bar{\bar{X}}_{h(k)} = \frac{1}{n''} \sum_{i=1}^{n''} X_{h(k)i}$$ = Mean for kth character on hth occasion based on units which are exclusively studied on hth occasion only.

$$\bar{X}'_{h(k)} = \frac{1}{n'} \sum_{i=1}^{m'} X_{h(k)i}$$ = Mean for kth character on hth occasion based on units which are common to two occasions.

$\bar{X}_{h(k)}$ = Sample mean for kth character on hth occasion based on all the n units.

We also define

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

$$\text{Cov}(X_{h(k)}, X_{r(s)}) = S_{h(k)r(s)} = \frac{1}{N-1} \sum_{i=1}^{N} (X_{h(k)i} - \bar{X}_{h(k)}) \times (X_{r(s)i} - \bar{X}_{r(s)}) ; \; r,s = 1,2.$$

And the correlation coefficient between sth character on rth occasion and kth character on hth occasion is given by

$$\rho_{h(k)r(s)} = \frac{s_{h(k)r(s)}}{s_{h(k)}\, s_{r(s)}}$$

It should be noted that no restriction have been imposed on the correlation coefficient or the population. Now efficient estimate of population means of both the two characters on the 1st occasion are given by simple means $\bar{\bar{X}}_{1(1)}$ & $\bar{\bar{X}}_{1(2)}$ based on n units.

Now an efficient estimate of population mean for 1st character on second occasion, utilizing information available for only corresponding 1st character on 1st occasion was given earlier.

$$_1\hat{\bar{\bar{X}}}_{2(1)} = a_1(\bar{X}'_{1(1)} - \bar{X}''_{1(1)}) + a_2\bar{X}'_{2(1)} + (1-a_2)\bar{X}''_{2(1)} \qquad \ldots\ldots(3.1)$$

But if the information available for 2nd character on 1st occasion is also used to improve the estimate then it is given by

$$_2\hat{\bar{\bar{X}}}_{2(1)} = a_1(\bar{X}'_{1(1)} - \bar{X}''_{1(1)}) + a_2(\bar{X}'_{1(2)} - \bar{X}''_{1(2)})$$

$$+ a_3\bar{X}'_{2(1)} + (1-a_3)\bar{X}''_{2(1)} \qquad \ldots\ldots(3.2)$$

and if the information available for 2nd character on 2nd occasion is also used, the estimate is given by

$$_3\hat{\bar{X}}_{2(1)} = a_1(\bar{X}'''_{1(1)} - \bar{X}''_{1(1)}) + a_2(\bar{X}'_{1(2)} - \bar{X}''_{1(2)})$$

$$+ a_3(\bar{X}'''_{2(2)} - \bar{X}''_{2(2)}) + a_4\bar{X}'_{2(1)} + (1-a_4)\bar{X}''_{2(1)}$$

.....(3.3)

Now from theorem given in the appendix it is easily seen that estimate (3.3) is more efficient than estimate (3.2) which is superior to (3.1).

To find the variance.

When we study two characters on two occasions an unbiased estimate of population mean for 2nd character on second occasion can be written as

$$\hat{\bar{X}}_{2(2)} = I_1(\bar{X}'''_{1(1)} - \bar{X}''_{1(1)}) + I_2(\bar{X}'_{1(2)} - \bar{X}''_{1(2)})$$

$$+ I_3(\bar{X}'''_{2(1)} - \bar{X}''_{2(1)}) + I_4\bar{X}'_{2(2)} + I_5\bar{X}''_{2(2)}$$

Such that $I_4 + I_5 = 1$ ......(3.4)

and its variance is given by

$$\begin{bmatrix} I_1 & I_2 & I_3 & I_4 & I_5 \end{bmatrix} \begin{bmatrix} V_{11} & V_{12} & V_{13} & V_{14} & V_{15} \\ V_{21} & V_{22} & & & \\ & & \ddots & & \\ & & & & \\ V_{51} & V_{52} & & & V_{55} \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \\ I_3 \\ I_4 \\ I_5 \end{bmatrix}$$

Here $V_{ij} = V_{ji}$, the covariance between two expressions in (3.4) having coefficients I_i & I_j.

The above expression for variance can be written in a simple form as LAL'.

Where L is a row matrix $(I_1 \; I_2 \; I_3 \ldots I_5)$ L' its transpose and A is variance covariance matrix.

Now for obtaining the best linear unbiased estimate we should choose $I_{(i)}$'s in such a way that $V(\hat{\bar{X}}_{2(2)})$ is minimum subject to condition however that $I_4 + I_5 = 1$.

Now for this consider the function

$F = LAL' - 2\lambda(LE' - 1)$ where E is the row matrix (0 0 0 1 1) and is undetermined multiplier

Now for minimizing variance we should have

$\frac{\partial F}{\partial L} = 0$ i.e. $2AL' - 2\lambda E' = 0$

i.e. $AL' = \lambda E'$

Now let $P = \frac{L}{\lambda}$ i.e. $P = (P_1 \; P_2 \ldots P_5)$ $P_i = \frac{I_i}{\lambda}$

Then $AP' = E'$

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

But $I_4 + I_5 = 1$

hence $\lambda = \dfrac{1}{P_4 + P_5}$

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

Thus the best estimator for $\bar{X}_{2(2)}$ is given by (3.4) where $I_1, I_2, \ldots, I_5$ are given by

$$L' = \frac{1}{P_4 + P_5} \quad A^{-1} \quad E'$$

and the minimum variance given by

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

We have got expressions for all quantities except V_{ij}, to get the estimate of mean and its variance. The variance covariance matrix can be found as

$$A = \begin{vmatrix} \dfrac{s^2_{1(1)}}{npq} & \dfrac{s_{1(1)1(2)}}{npq} & \dfrac{s_{1(1)2(1)}}{np} & \dfrac{s_{1(1)2(2)}}{np} & 0 \\ & \dfrac{s^2_{1(2)}}{npq} & \dfrac{s_{1(2)2(1)}}{np} & \dfrac{s_{1(2)2(2)}}{np} & 0 \\ & & \dfrac{s^2_{2(1)}}{npq} & \dfrac{s_{2(1)2(2)}}{np} & \text{[illegible]} \\ & & & \dfrac{s^2_{2(2)}}{np} & 0 \\ & & & & \dfrac{s^2_{2(2)}}{nq} \end{vmatrix}$$

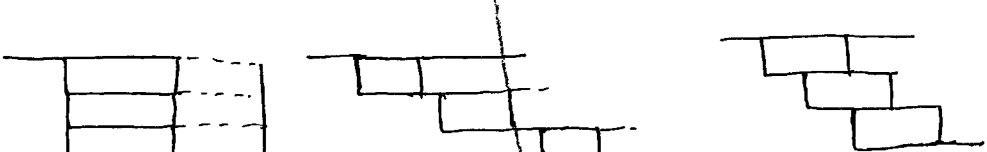
Sampling for k characters on h occasions.

When the study is to be continued for more than two occasions, the sample can be drawn in 3 different ways as

(1) Let n be the sample size on 1st occasion, now for all the successive occasions we keep a fixed s.r.s. of np units for all occasions from n units in the sample on 1st occasion and on each occasion we supplement it by an independent sample of nq units, selected from the (N-n) units which are not observed on the previous occasion, so that sample size on each occasion is kept fixed. In this way on some occasion after 2nd some units in the fresh sample on that occasion may come which have been studied on some previous occasion also, but here we assume that they provide no additional information.

2. When the sample of np units is retained from among the nq units which have been studied only on the current occasion i.e. no units are common between two occasions, one occasion apart.

3. When a simple random sample of np units is retained for the next occasion from n units observed on the previous occasion only.



The results in the thesis have been developed for the sampling pattern 1. But with minor changes and and alternations these can be easily extended to other schemes also.

Now the estimate of mean and its variance for general case of k characters studied on each of h occasions for sampling pattern 1 can be written very easily.

$$\bar{X}_{h(k)} = a_1 (\bar{X}'_{1(1)} - \bar{X}''_{1(1)}) + a_2 (\bar{X}'_{1(2)} - \bar{X}''_{1(2)}) \ldots\ldots$$

$$\ldots a_k (\bar{X}'_{1(k)} - \bar{X}''_{1(k)}) + a_{k+1} (\bar{X}'_{2(1)} - \bar{X}''_{2(1)}) + a_{k+2} (\bar{X}'_{2(2)} - \bar{X}''_{2(2)})$$

$$+ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$+ a_{(h-1)k+1} (\bar{X}'_{h(1)} - \bar{X}''_{h(1)}) + \ldots\ldots\ldots\ldots\ldots\ldots$$

$$+ a_{hk} \bar{X}'_{h(k)} + a_{hk+1} \bar{X}''_{h(k)}$$

where

$$a_{hk} + a_{hk+1} = 1.$$

and $V(\bar{X}_{h(k)}) = LVL'$

where L is row matrix $(a_1, a_2, \ldots\ldots a_{hk+1})$

and L' is its transpose and V is variance covariance matrix as defined earlier.

Now as proved earlier this estimate will be minimum variance estimate when

$$L' = \frac{1}{P_{hk} + P_{hk+1}} \quad A^{-1} E'$$

where E is row matrix $(0, 0, \ldots\ldots 1, 1)$ of order hk+1 and $P = (p_1, p_2, \ldots\ldots\ldots p_{hk+1})$

where $p_i = \dfrac{a_i}{\lambda}$

and since $a_{hk} + a_{hk+1} = 1$

and so $\lambda = \dfrac{1}{p_{hk} + p_{hk+1}}$

And the minimum variance is given by

$$LAL' = LE' = \lambda^2 \; E \, A^{-1} E'$$

$$= \left(\frac{1}{p_{hk}+p_{hk+1}}\right)^2 E \, A^{-1} E'$$

Now since all the quantities except A are known so we can easily get a minimum variance unbiased estimate if only A is known. For our study A is given below.

$$A = \begin{vmatrix} \frac{A^2_{1(1)}}{npq} & \frac{A_{1(1)1(2)}}{npq} \cdots \frac{A_{1(1)1(k)}}{npq} & \frac{A_{1(2)2(1)}}{np} \cdots & \frac{A_{1(2)h(k)}}{np} & \frac{A_{1(1)h(k)}}{np} & 0 \\ & \frac{A^2_{1(2)}}{npq} \cdots \frac{A_{1(2)1(k)}}{npq} & \frac{A_{1(2)2(1)}}{np} \cdots\cdots\cdots\cdots & & \frac{A_{1(2)h(k)}}{np} & 0 \\ & & & & \frac{A^2_{h(k)}}{np} & 0 \\ & & & & & \frac{A^2_{h(k)}}{nq} \end{vmatrix}$$

CHAPTER - IV

Sampling on h occasions for k characters for a two stage design.

In this chapter we develope the formulaes of previous chapter for two stage design.

Let N and M be the no. of p.s.u and s.s.u's in the population and suppose that a simple random sample of n p.s.u's is drawn for the first occasion and that from each p.s.u. selected we select m s.s.u's so that sample on the 1st occasion consists of nm sampling units.

Now suppose that we retain a s.r.s. of np p.s.u's for the sample on second occasion and suppliment it by nq independently selected p.s.u's where p + q = 1, so that no. of p.s. units on second occasion is also n. Now in the p.s.u's retained from previous sample we retain all s.s.u's which were observed on 1st occasion and in the fresh nq p.s.u's also we select m s.s.u's from each of nq p.s.u so that total sample size on 2nd occasion is again npm + nqm = nm(p+q)=nm.

Practical considerations in an actual field survey necessitate the size of sample to be the same on all occasion.

Now when the survey is continued for more than two occasions, the sample can be drawn in many different ways but we shall consider the case when fixed s.r.s. of npm units from the nm units on 1st occasion is observed over all the h occasions and on each occasion the sample is supplemented by independently selected nqm units from (N-n)m units which were not observed on the previous occasion so that sample size remains fixed on each occasion.

We assume throughout the investigation that N and M are large enough to neglect

$\frac{1}{N}$ & $\frac{1}{M}$.

Now we shall denote by $X_{h(k)ij}$ the value of jth s.s.u in the ith p.s.u for kth character on hth occasion.

We shall have two components in the sample on each occasion. Their means will give the estimates for the corresponding population means.

We define these estimates as follows

$$\overline{X}'_{h(k)} = \frac{1}{npm} \sum_{i}^{np} \sum_{j}^{m} X_{h(k)ij}$$

= Mean on hth occasion for kth character based on npm units which are common to all the occasions.

$$\overline{X}''_{h(k)} = \frac{1}{nqm} \sum_{i}^{nq} \sum_{j}^{m} X_{h(k)ij}$$

= Mean on hth occasion for kth character based on nqm units which are selected afresh.

and $\bar{X}_{h(k)}$ the sample mean based on all the nm units in the sample.

Now we define

$$S^2_{h(k)i} = \frac{1}{M-1} \sum^{M} (X_{h(k)ij} - \bar{X}_{h(k)i})^2$$

$$\bar{S}^2_{h(k)w} = \frac{1}{N} \frac{1}{M-1} \sum^{N} \sum^{M} (X_{h(k)ij} - \bar{X}_{h(k)i})^2$$

$$= \frac{1}{N} \sum^{N} S^2_{h(k)i}, \text{ where } \bar{X}_{h(k)i} = \frac{1}{M} \sum_{j}^{M} X_{h(k)ij}$$

Similarly

$$S_{h(k)r(s)w} = \frac{1}{N} \frac{1}{M-1} \sum^{N} \sum^{M} (X_{h(k)ij} - \bar{X}_{h(k)i}) \times (X_{r(s)ij} - \bar{X}_{r(s)i})$$

Also

$$S_{h(k)b} = \frac{1}{N-1} \sum^{N} (\bar{X}_{h(k)i} - \bar{X}_{h(k)})^2$$

and

$$S_{h(k)r(s)b} = \frac{1}{N-1} \sum^{N} (\bar{X}_{h(k)i} - \bar{X}_{h(k)})(\bar{X}_{r(s)i} - \bar{X}_{r(s)})$$

$$\rho_{h(k)r(s)1} = \frac{S_{h(k)r(s)b}}{S_{h(k)b}\, S_{r(s)b}}$$

$$\rho_{h(k)r(s)2} = \frac{S_{h(k)r(s)w}}{\bar{S}_{h(k)w}\, \bar{S}_{r(s)w}}$$

Estimates of these quantities from the sample are given by

$$\text{Est. } S^2_{h(k)b} = s^2_{h(k)} - \frac{\bar{s}^2_{h(k)w}}{m}$$

$$\text{Est. } \bar{S}^2_{h(k)w} = \bar{s}^2_{h(k)w} = \frac{1}{n} \sum_{i=1}^{n} s^2_{h(k)i}$$

where

$$s^2_{h(k)b} = \frac{1}{np-1} \sum^{np} (\bar{X}_{h(k)i} - \bar{\bar{X}}_{h(k)})^2$$

~~$s^2_{h(k)w}$~~

$$s_{h(k)r(s)b} = \frac{1}{np-1} \sum^{np} (\bar{X}_{h(k)i} - \bar{\bar{X}}_{h(k)})(\bar{X}_{r(s)i} - \bar{\bar{X}}_{r(s)})$$

and

$$\bar{s}^2_{h(k)w} = \frac{1}{np} \frac{1}{(m-1)} \sum^{np} \sum^{m} (X_{h(k)ij} - \bar{X}_{h(k)i})^2$$

$$s_{h(k)r(s)w} = \frac{1}{np} \frac{1}{(m-1)} \sum^{np} \sum^{m} (X_{h(k)ij} - \bar{X}_{h(k)i}) \times$$

$$(X_{r(s)ij} - \bar{X}_{r(s)i})$$

Now

$$\hat{V}(\bar{\bar{X}}'_{h(k)}) = \frac{1}{np} \left\{ s^2_{h(k)b} + \frac{\bar{s}^2_{h(k)w}}{m} \right\} = \frac{A^2_{h(k)}}{np}$$

$$\hat{Cov}(\bar{\bar{X}}'_{h(k)} \; \bar{\bar{X}}'_{r(s)}) = \frac{1}{np} \left\{ s_{h(k)r(s)b} + \frac{\bar{s}_{h(k)r(s)w}}{m} \right\}$$

$$= \frac{A_{h(k)r(s)}}{np}$$

$$\hat{V}(\bar{\bar{X}}''_{h(k)}) = \frac{1}{nq} \left\{ (s^2_{h(k)b}) + \frac{\bar{s}^2_{h(k)w}}{m} \right\} = \frac{A^2_{h(k)}}{nq}$$

Now an unbiased estimate of mean for kth character on hth occasion and its variance can be obtained just on the similar lines as in the single stage design.

CHAPTER - V

Application of the results in Agronomic and Agro-economic surveys in the Intensive Agricultural District Programme District Aligarh.

The I.A.D.P. envasages the selection of favourable areas with maximum irrigation facilities and providing all the elements such as full supplies, credit, etc. to increase Agricultural Production.

An integral part of I.A.D.P. is provision for analysis and evaluation of programme from its inception. The bench-mark and assessment surveys are organised in I.A.D.P. districts with the object of assessing the changes being brought about in agronomic practices by cultivators and the consequent improvement in their yields. The field work relating to surveys consists of two components.

(a) Agronomic and Agro-economic survey

(b) Crop-cutting experiments on important crops.

For our utilization we have studied here component (a) Agronomic and Agro-economic surveys.

Sampling plan.

The sampling plan is suggested by I.A.R.S. and is based on stratified multistage random sampling technique. A zone consisting of 2 to 4 blocks constitutes a stratum. A village and a cultivators holding in the village are 1st and 2nd stage sampling units respectively.

From the whole population about 100 villages are selected and from each selected village 8 cultivators holding are convassed each year. Sampled cultivators in a total no. of 24 villages selected on 1st occasion have been kept fixed for convassing from year to year. In addition a fresh sample of about 76 villages is selected each year from those villages which are not selected on the previous occasion and from each selected village 8 cultivators are convassed so that sample size on each occasion remains the same.

The survey was started in the year 1962-63 and the fixed villages were changed in 1967-68 the data for the year 1962-63 was not available in the compact and required form. So the data for the 4 years from 1963-64 to 1966-67 is utilized for the present investigation.

The main character under study has been taken to be consumption of nitrogeneous fertilizer per cultivator in the year 1966-67 with the help of auxiliary variaties available (a) Total irrigated area (b) Total area under all crops, on all the previous occasions as well as on the current occasion. Different estimates utilizing different no. of auxilary variates have been considered and their efficiency is calculated. The results are given in table (1).

We have considered the following different estimates.

$\bar{Y}_1$ Mean consumption of nitrogeneous fertilizer per cultivator for the year 1966-67 utilizing the information for the characters (a) Total irrigation area (b) Total area under all crops for the year 1966-67 and for all the three characters on all previous occasions.

$\bar{Y}_2$ Mean consumption of nitrogeneous fertilizer on all the three above characters on all the three previous occasions only.

$\bar{Y}_3$ Mean utilizing information on previous 2 occasions only on all the three characters.

$\bar{Y}_4$ Mean utilizing information on previous two occasions only for the character under study.

$\bar{Y}_5$ Simple mea-n of charater under study only on current occasion.

<u>Table 1-1</u>

Estimate	Mean	Variance
$\bar{Y}_1$	81.9701	36.8815
$\bar{Y}_2$	76.0350	49.3715
$\bar{Y}_3$	89.7037	51.5105
$\bar{Y}_4$	93.6563	57.6304
$\bar{Y}_5$	87.5325	63.7075

Now from the results obtained it is easily seen that as the utilization of auxiliary information is increased, the efficiency of the estimate increases and thus in multipurpose surveys the information on other correlated characters on previous occasion can be utilized to get better estimates of any character.

SUMMARY AND CONCLUSIONS

Expressions for an efficient estimate of population mean for a character and its variance on 2nd occasion have been obtained utilizing information on the same character on the 1st occasion. The sample estimate of population variance is not an unbiased one and the expression for the relative bias in the estimate of variance has been given in Chapter II. It has been proved that relative bias is a monotonically decreasing function of n the sample size.

In Chapter III the expression for the estimate of mean and its variance of character on second occasion have been obtained utilizing information on another correlated character on the second occasion and also the information for both character on 1st occasion. Then the results have been extended to the general case of studying information on k characters on h occasions, $h \geq 2$, $k \geq 2$ and an efficient estimate of mean for kth character on hth occasion and its variance have been obtained.

In Chapter IV the results have been extended to a two stage design when replacement is done only in p.s.u.s. ~~applied of results~~.

In Chapter V the results has been illustrated by the data obtained in the Agronomic and Agro-economic surveys of the I.A.D.P. District Aligarh. By considering the consumption of nitrogeneous fertilizer per cultivator it has been observed that as we increase the use of auxiliary variates the estimates become more and more efficient. The comparisons have been made of efficiencies of different estimate having different number of auxiliary vanates and different number of occasions on which the survey has been conducted.

APPENDIX

Theorem Let $\bar{\bar{Y}}_{lrq}$ and $\bar{\bar{Y}}_{lrk}$ denote respectively the unbiased regression estimators of $\bar{\bar{Y}}$ based on the set of supplementary variables $(X_1), (X_2).....(X_q)$ and $(X_1), (X_2).....(X_k)$ where k is greater than q then

$$V(\bar{\bar{Y}}_{lrq}) \geq V(\bar{\bar{Y}}_{lrk})$$

The proof of the theorem is similar to Olkin (1958) and hence is ~~submitted~~ Ommitted here.

R E F E R E N C E S

1. Jesson, R.J. (1942) Statistical investigation of a sample survey for obtaining farm facts, IOWA Agri. Expt. Stat. Res. Bull. 304.

2. Yates F. (1949) Sampling methods for censuses and surveys, Charles Griffin and Co., London.

3. Patterson, H.D. (1950) Sampling on successive occasion with partial replacement of units J.R. SS series B-12.

4. Sukhatme, P.V. (1953) Sampling theory of surveys with applications, Ind. Soc. of Agri. Stat. New Delhi. The Iowa State College Press Ames, Iowa, U.S.A.

5. Tikkiwal B.D. (1954) Theory of successive multiphase sampling abstract Ann. Math. Stat.25,174.

6. Tikkiwal, B.D. (1956) On efficiency of estimates in successive multiphase sampling, Abstract Ann. Math. Stat.26: 158.

7. Hansen & Hurwitz (1956) Sample surveys, Method and theory, JohnWiley & Sons, New York.

8. Singh, D. (1968) Estimates in successive sampling using multistage Design Jr. of American Stat. Association.

9. Singh & Kathuria (1968) "On two stage successive sampling" Australian Jr. of Statistics. Vol. 11 No. 2.

10. Cochran, W.G. (1963)"Sampling techniques" 2nd edition, John Wiley & Sons.

11. Talukdar, A.H. (1968) Multivariate regression, estimates for finite population. Ann of the Institute of Statistical Mathematics V. 20, No.3.

12. Olkin (1958) Multivariate ratio estimation for finite populations Biometrica 45, 1954-165.

13. Parbhu Ajgachkar, S.C. & Tikkiwal B.D., (1961) on the theory of unvariate successive sampling Ann. Math. Stat.32, 340 (Abstract).

14. Edeler, A.R. (1955) Rotation sampling Ann. Math. Stat. 26, 664-685.

15. Tikkiwal B.D. (1964) A note on two stage successive occasions sankhya, 26, 97-100.

16. Tikkiwal B.D. (1958) An examination of the effect of method sampling on the efficiency of estimation in the theory of successive sampling Jr. Ind. Soc. Agri. Stat.10, 16-21.

17. Tikkiwal B.D. (1965) The theory of two stage sampling on successive occasion, Jr. of Indian Stat. Association Vol. III No. I 125-135.