V84 V95

OPTIMUM ALLOCATION IN MULTIPHASE OR SUCCESSIVE SAMPLING INVESTIGATIONS WITH TWO-STAGE SAMPLING

by

M.G. Mittal

Dissertation submitted in fulfilment of the requirements for the award of Diploma in Agricultural and Animal Hubbandary Statistics of the Institute of Agricultural Research Statistics(ICAR), New Delhi

1966

ACKNOWLEDGEMENTS

I have great pleasure in expressing my deepest sense of gratitude to Dr G.R. Seth, Statistical Adviser, Indian Council of Agricultural Research for his valuable guidance, keen interest and constant encouragement throughout the course of investigation and of the preparation of the thesis. I am also deeply indebted to him for providing me adequate facilities during the course of investigation of the thesis.

M. G. Mittal

(M. G. Mittal)

C_O_N_T_E_N_T_S

INTRODUCTION

The following problems have been attempted in the present work.

I. Investigation of optimum design in Multi-stage sampling when the survey is to be conducted on two successive occasions.

II. To determine the estimation procedure and investigation of the optimum design in Three Phase Sampling.

CHAPTER I

Matched sampling on two occasions, with two-stage sampling.

1.1. Introduction.

Usually the results of a sample survey are valid for the reference period for which data has been collected in the survey. Such results remain valid for future periods as long as the pattern of variation in the population does not change its character. But, for a dynamic population, that is, the one, which is subject to change from time to time, such as the extent of area under improved seeds, the extent of fertilizer use, or the number of unemployed persons in a country, a single survey conducted for a certain refeerence period is of limited use unless it is repeated frequently at regular intervals of time to bring the information up-to-date. In case of such repeated surveys a question arises regarding the retention in later rounds of the units observed in earlier rounds, that is, whether all or a fraction of these units should be retained, or all replaced in successive rounds. For studying the change in the population character it will be advisable to retain all the units from the previous round. On the other hand, if the objective is to study the average or the total of the character over a number of rounds, it will be better to change all the units. This is on the assumption that the character under study on any two occasions is positively correlated. But, usually we are interested in the study of the change as well as the population value on each round, and also the total or the average over a number of rounds. For such situations a partially replaced sample may be more efficient. Partial or full retention of the sample units observed on each occasion cuts down on the cost of the surveys as there is a saving in the cost of selection of the units as well as considerable saving in time required for establishing contacts. On the other hand, in some cases repeated recording of the same units over successive occasions may cause difficulties in getting the cooperation of the respondents.

Several authors such as Uessen (1942), Yates(1949), Patterson (1950), Tikkiwal (1951), (1954), (1955), (1956), (1958), (1960), Nardin (1953), Hansen, Hurwitz and Madow (1953), Eckler (1955), Hansen (1955), B.D. Singh (1962), Seshavadhanulu (1963), Savdasia (1964) have developed efficient estimates for cases involving unistage sampling for successive occasions.

Singh D. (1959), has studied the problem of partial replacement of primary units involving two-stage sampling for two occasions only. He has obtained the expression for the estimate of the population mean and variance of the estimate.

Kathuria (1959) has extended the results of D. Singh(1959) to more than two occasions. He has considered the cases when (i) the replacement is done for primary stage units only and (ii) when the replacement is done for second stage units only. He has also obtained the optimum value of replacement fraction without taking into consideration the cost aspect. Further, for a given replacement fraction, he has found optimum allocation of primary stage units and second stage units taking into consideration a suitable cost function.

In the present work, an attempt has been made to obtain expression for optimum allocation and sampling fraction for a two-stage sampling design when the survey is to be conducted on two successive occasions and cost to be incurred does not change from the one occasion to another. For the sake of simplicity, the discussion is confined to a two-stage design where all the sampling units at both the stages were drawn with equal probability and without replacement and out of the primary stage units examined at the first occasion a fraction p was retained for the second occasion.

1.2. Notation and Estimation Procedure.

Let

N = Number of primary stage units in the population, which will be assumed to be large.

M = Number of secondary stage units in each primary stage unit.

n = Number of primary units in the sample on each occasion.

m = Number of secondary units to be selected from each primary unit in the sample on each occasion.

p = Proportion of primary units to be retained for the second occasion.

$q = 1-p$

np is assumed to be an integer.

$X_{ij}^{(U)}$ = Value of the character X of the jth secondary unit of the ith first stage unit on the Uth occasion.

$U = 1,2$

$j = 1,2.........m$

$i = 1,2.........n$ for $U = 1$

$i = 1,2,\ldots\ldots np,$

$(n+1,\ldots\ldots n+qn)$ for $U = 2$

$\bar{X}^{(U)}_{i(m)}$, $\bar{X}^{(U)}_{i(M)}$ relates to the sample and population means of ith primary unit on the Uth occasion.

$\bar{X}^{(1)}_{(np,m)}$ = mean of the values of the character under study on the first occasion for the mnp secondary units which are common to the second occasion.

$\bar{X}^{(2)}_{(np,m)}$ = mean of the values of the character on the second occasion of mnp secondary units common with the first occasion.

$\bar{X}^{(1)}_{(nq,m)}$ = mean on the first occasion based on mnq units which are not retained on the second occasion.

$\bar{X}^{(2)}_{(nq,m)}$ = mean on the second occasion based on mnq units which are selected afresh.

$$\bar{X}^{(U)}_{(n,m)} = 1/n \sum_{i=1}^{n} \bar{X}^{(U)}_{i(m)} \qquad (U = 1,2)$$

$\bar{X}^{(U)}_{N,M}$ = population mean for ultimate unit on the Uth occasion. $(U = 1,2)$.

Let us consider the following estimate.

$$\hat{\bar{X}}^{(2)}_{N,M} = \left\{ \bar{X}^{(2)}_{(np,m)} + b \left(\bar{X}^{(1)}_{(n,m)} - \bar{X}^{(1)}_{(np,m)} \right) \right\} (1 - \phi^{(2)}) + \phi^{(2)} \bar{X}^{(2)}_{(nq,m)}$$

where 1.2.1

$$b = \frac{\hat{Cov}\left(\bar{X}^{(2)}_{(np,m)} , \bar{X}^{(1)}_{(np,m)} \right)}{q\left\{ \hat{V}\left(\bar{X}^{(1)}_{(np,m)}\right) + \hat{V}\left(\bar{X}^{(1)}_{(nq,m)}\right) \right\}} \qquad \ldots\ldots 1.2.2$$

Neglecting* sampling error in b,

variance of $(\hat{\bar{X}}^{(2)}_{N,M})$ is minimum when $\phi^{(2)}$ is given by

$$\phi^{(2)} = \frac{V\left\{\bar{X}^{(2)}_{(np,m)} + b\left(\bar{X}^{(1)}_{(n,m)} - \bar{X}^{(1)}_{(np,m)}\right)\right\}}{V\left\{\bar{X}^{(2)}_{(np,m)} + b\left(\bar{X}^{(1)}_{(n,m)} - \bar{X}^{(1)}_{(np,m)}\right)\right\} + V\left(\bar{X}^{(2)}_{(nq,m)}\right)} \qquad \ldots 1.2.3$$

Following Sukhatme(1953)

$$V\left(\bar{X}^{(U)}_{(np,m)}\right) = \frac{1}{np}\left(S_b^{(U)}\right)^2 + \frac{1}{npm}\,\frac{1}{N}\sum_{i=1}^{N}\left(S_i^{(U)}\right)^2 \qquad \ldots 1.2.4$$

assuming M,N to be large.

Where

$$\left(S_b^{(U)}\right)^2 = 1/N \sum_{i=1}^{N}\left(\bar{X}^{(U)}_{i(M)} - \bar{X}^{(U)}_{N,M}\right)^2$$

$$\left(S_i^{(U)}\right)^2 = 1/M \sum_{j=1}^{M}\left(X^{(U)}_{ij} - \bar{X}^{(U)}_{i(M)}\right)^2 \qquad U = 1,2$$

From (1.2.4) we have

$$V\left(\bar{X}^{(U)}_{(np,m)}\right) = \frac{\left(S_b^{(U)}\right)^2}{np} + \frac{\left(\bar{S}_w^{(U)}\right)^2}{mnp}$$

$$= \frac{\alpha'^{(U)}}{np} \qquad \text{say} \qquad \ldots 1.2.5$$

---

* neglecting sampling error in 'b' the bias involved in the $V(\bar{X}^{(2)}_{N,M})$ is negligible as shown in Appendix I.

Where $\alpha'^{(U)} = (S_b^{(U)})^2 + \frac{(\bar{S}_w^{(U)})^2}{m}$

and $(\bar{S}_w^{(U)})^2 = 1/N \sum_{i=1}^{N} (\bar{S}_i^{(U)})^2$ .... U = 1,2

Similarly

$$V(\bar{X}^{(U)}_{(nq,m)}) = \frac{\alpha'^{(U)}}{nq}$$ 1.2.6

....U = 1,2

$$Cov\left\{ \bar{X}^{(1)}_{(np,m)}, \bar{X}^{(2)}_{(np,m)} \right\} = \frac{S^{(1)(2)}}{np} + \frac{\bar{S}_w^{(1)(2)}}{mnp} = \frac{\gamma}{np}$$ 1.2.7

Where

$$\gamma = S_b^{(1)(2)} + \frac{\bar{S}_w^{(1)(2)}}{m}$$

$$S_b^{(1)(2)} = 1/N \sum_{i=1}^{N} (\bar{X}^{(2)}_{i(M)} - \bar{X}^{(2)}_{N,M})(\bar{X}^{(1)}_{i(M)} - \bar{X}^{(1)}_{N,M})$$

$$S_i^{(1)(2)} = \frac{1}{M} \sum_{j=1}^{M} (X^{(2)}_{ij} - \bar{X}^{(2)}_{i(M)})(X^{(1)}_{ij} - \bar{X}^{(1)}_{i(M)})$$

and $\bar{S}_w^{(1)(2)} = \frac{1}{N} \sum_{i=1}^{N} S_i^{(1)(2)}$

If we assume

$(S_b^{(2)})^2 = (S_b^{(1)})^2 = S_b^2$ say

$(\bar{S}_w^{(2)})^2 = (\bar{S}_w^{(1)})^2 = \bar{S}_w^2$ say i.e. $\alpha'^{(U)} = \alpha$ say $\qquad U = 1,2$

Then after simplifications we get

$$V(\hat{\bar{X}}^{(2)}_{N,M}) = \frac{\alpha(\alpha^2 - \gamma^2 q)}{n(\alpha^2 - \gamma^2 q^2)} \qquad \ldots 1.2.8$$

Where $\alpha = S_b^2 + \frac{\bar{S}_w^{(2)}}{m}$

Let us assume that

$$\rho_i^{(1)(2)} = \rho_w^{(1)(2)} = \rho_w \quad \text{say}$$

where $\rho_i^{(1)(2)}$ is the correlation coefficient between the values on the first and second occasion of secondary units within ith primary unit.

Then

$$\bar{S}_w^{(1)(2)} = \rho_w \bar{S}_w^2$$

Similarly

$$S_b^{(1)(2)} = \rho_b S_b^2$$

Where

$$\rho_b = \frac{\sum_{i=1}^{N} (\bar{X}^{(2)}_{i(M)} - \bar{X}^{(2)}_{N,M})(\bar{X}^{(1)}_{i(M)} - \bar{X}^{(1)}_{N,M})}{\sqrt{\left\{\sum_{i=1}^{N} (\bar{X}^{(2)}_{i(M)} - \bar{X}^{(2)}_{N,M})^2\right\}\left\{\sum_{i=1}^{N} (\bar{X}^{(1)}_{i(M)} - \bar{X}^{(1)}_{N,M})^2\right\}}} \qquad \ldots 1.2.9$$

$$\rho_w = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M}(X_{ij}^{(2)} - \bar{X}_{i(M)}^{(2)})(X_{ij}^{(1)} - \bar{X}_{i(M)}^{(1)})}{\sqrt{\left\{\sum_{i=1}^{N}\sum_{j=1}^{M}(X_{ij}^{(2)} - \bar{X}_{i(M)}^{(2)})^2\right\}\left\{\sum_{i=1}^{N}\sum_{j=1}^{M}(X_{ij}^{(1)} - \bar{X}_{i(M)}^{(1)})^2\right\}}} \quad \ldots 1.2.10$$

For further algebraic simplifications we shall make use of the following notations.

$B_1 = S_b^{(1)(2)}$, $A_1 = S_b^2$

$B_2 = \bar{S}_w^{(1)(2)}$, $A_2 = \bar{S}_w^2$

$K = \frac{\gamma}{\alpha}$

Expression (1.2.8) can now be written as

$$V(\hat{\bar{X}}_{N,M}^{(2)}) = \frac{\alpha(1-K^2 q)}{m(1-K^2 q^2)} \quad \ldots 1.2.11$$

1.3 To determine optimum values of m.n and q.

Let us assume that the total cost on both the occasions can be represented by

$$C = C_1 n + C_2 mn + C_1 qn + C_2' mnp + C_2 mnq \quad \ldots 1.3.1$$

Where

C_1 = Cost of listing per primary unit.

C_2 = Cost of enumerating a secondary unit for the first time.

C_2' = Cost of enumerating on the second occasion a secondary stage unit which has already been observed on the first occasion.

C_2' is assumed to be less than or equal to C_2.

In the above cost function we have also assumed that on the second occasion no re-listing of primary units once listed is required.

Now we have to choose q, n and m such as to minimise the variance of $(\hat{\bar{X}}^{(2)}_{N,M})$ *subject to a fixed cost C.*

Putting the value of n from (1.3.1) in (1.2.11) we have

$$V(\hat{\bar{X}}^{(2)}_{N,M}) = \frac{C_1 + (C_2 + C_2')\ m + (C_1 + C_2 m - C_2' m)q}{C} \times \frac{\alpha(1-K^2 q)}{1-K^2 q^2} \qquad \ldots\ldots 1.3.2$$

Minimising this with respect to m and q we obtain

$$\frac{\partial V}{\partial m} = 0 \implies l_4 q^4 + l_3 q^3 + l_2 q^2 + l_1 q + l_0 = 0 \qquad \ldots\ldots 1.3.3$$

Where l_j (j=0,1,2,3,4) are polynomials of sixth degree in 'm' with coefficients depending upon

$$\frac{C_1}{C_2},\ \frac{C_2'}{C_2} \text{ and } \bar{S}_w^2 / S_b^2.$$

$$\frac{\partial V}{\partial q} = 0 \implies K^2 q^2 \left\{ (C_1 + C_2 m - C_2' m) - K^2 (C_1 + C_2 m + C_2' m) \right\} +$$

$$4K^2 q C_2' m + \left\{ (C_1 + C_2 m - C_2' m) - K^2 (C_1 + C_2 m + C_2' m) \right\}$$

$$= 0 \qquad \ldots 1.3.4$$

If $K^2 = \frac{C_1 + C_2 m - C_2' m}{C_1 + C_2 m + C_2' m}$, then $q = 0$.

then this provides value of m and for optimum it has to be consistent with the solution of the equation $l_0 = 0$

If $K^2 > \frac{C_1 + C_2 m - C_2' m}{C_1 + C_2 m + C_2' m}$

solving the equation 1.3.4, we get

$$q = \frac{2KC_2' m \pm \sqrt{(1-K^2)\left\{K(C_1 + C_2 m + C_2' m) - (C_1 + C_2 m - C_2' m)\right\}\left\{K(C_1 + C_2 m + C_2' m) + (C_1 + C_2 m - C_2' m)\right\}}}{K\left\{K^2 (C_1 + C_2 m + C_2' m) - (C_1 + C_2 m - C_2' m)\right\}}$$

$$\ldots 1.3.5$$

It can be easily seen that q is real if and only if $|K|$ lies between $\frac{C_1 + C_2 m - C_2' m}{C_1 + C_2 m + C_2' m}$ and 1. $\qquad \ldots 1.3.6$

Further it may be seen that roots when real are either both +ve and both -ve. In case the roots are positive, we must have

$$K^2 > \frac{C_1 + C_2 m - C_2' m}{C_1 + C_2 m + C_2' m} \qquad \ldots 1.3.7$$

$$|K| > \frac{C_1 + C_2 m - C_2' m}{C_1 + C_2 m + C_2' m}$$

From 1.3.6 and 1.3.7, it is clear that (1.3.4) has real +ve roots provided

$$1 > K^2 > \frac{C_1 + C_2 m - C_2' m}{C_1 + C_2 m + C_2' m} \quad ..(A)$$

$$1 \quad |K| > \sqrt{\frac{C_1 + C_2 m - C_2' m}{C_1 + C_2 m + C_2' m}}$$

After eliminating q from 1.3.3 and 1.3.5 we get an equation of degree 44 in 'm' which is to be solved under restriction (A). The equation of degree 44 in 'm' depends ~~depends~~ upon the 5 parameters $\frac{C_1}{C_2}$, $\frac{C_2'}{C_2}$, $\frac{\bar{S}_w^2}{S_b^2}$, ρ_b and ρ_w. For any given set of values for these parameters, this equation can be solved with the help of electronic computer. In case this equation has only one +ve real root, the substitution of this value in equation(1.3.5) will give the value of q.

If there are more than one positive real roots of the above equation; we shall only consider those values of m for which q lies between 0 and 1 and least value of m among those m, which provides the estimate with minimum variance. After determining the value of 'q' and 'm' we will substitute these values in equation (1.3.1) and determine 'n'.

Thus we get optimum value of q, m and n.

Numerical solution of this equation is complicated and would require considerable time even on a computor like 1620 IBM. This equation reduces to a somewhat simpler equation if we assume some relationship between C_1, C_2, ρ_b and ρ_w. One simple case is discussed in the following section.

1.4 Case when $\rho_b = \rho_w =$ (say) and $C_2' = C_2$

In this case from (1.2.11) we have

$$V(\hat{\bar{X}}^{(2)}_{N,M}) = \frac{\alpha(1-\rho^2 q)}{n(1-\rho^2 q^2)}$$

$$\therefore K = \frac{B_1 + B_2/M}{A_1 + A_2/m} = \frac{\rho_b A_1 + \rho_w A_2/m}{A_1 + A_2/m}$$

$$= \frac{A_1 + A_2/m}{A_1 + A_2/m}$$

$$= \rho$$

and (1.3.1) becomes

$$C = C_1 n + 2C_2 mn + C_1 q n \qquad \ldots.1.4.2$$

putting the value of n from (1.4.2) in (1.4.1) we have

$$V(\hat{\bar{X}}^{(2)}_{N,M}) = \frac{C_1 + 2C_2 m + C_1 q}{C} \times \frac{\alpha(1-\rho^2 q)}{(1-\rho^2 q^2)} \qquad \ldots.1.4.3$$

minimising this with respect to m and q we obtain

$$\frac{\partial V}{\partial q} = 0 \Longrightarrow \rho^2 q^2 \left\{ \lambda(1-\rho^2) - 2m\rho^2 \right\} + 4m\rho^2 q$$

$$+ \left\{ \lambda(1-\rho^2) - 2m\,\rho^2 \right\} = 0 \qquad \ldots.1.4.4$$

$$\frac{\partial V}{\partial m} = 0 \Longrightarrow 2m^2 - \varepsilon\phi - \varepsilon\phi\, q = 0$$

$$\text{or } q = \frac{2m^2}{\varepsilon\phi} - 1$$

$$= \frac{2m^2 - \varepsilon\phi}{\varepsilon\phi} \qquad \ldots\ldots 1.4.5$$

Putting the value of q from equation (1.4.5) in equation (1.4.4) we get

$$8m^5\rho^4 - m^4\left\{4\rho^2\varepsilon(1-\rho^2)\right\} - m^3\left\{8\varepsilon\phi\rho^2(1-\rho^2)\right\}$$

$$+ m^2\left\{4\varepsilon^2\phi^2(1-\rho^2)\rho^2\right\} + m\left\{2\varepsilon^2\phi^2\rho^2(\rho^2+3)\right\} - \varepsilon^3\phi^2(1-\rho^4) = 0$$

$$\ldots\ldots 1.4.6$$

This equation can be solved easily. We shall find that root for which q lies between 0 and 1 after substituting the value in equation (1.4.5).

Then putting these values of m and q we can determine n from equation(1.4.2).

Thus we obtain the optimum value of m, n and q.

Equation (1.4.6) has at most three real roots. As the product of the root is +ve, there is at best one real +ve root. In case of only one +ve root, it is meaningful for our case provided it lies between $\frac{\varepsilon\phi}{2}$ and $\varepsilon\phi$ i.e. if the equation by substituting $m = t\sqrt{\varepsilon\phi}$, the equation $F(t) = 0$ $\ldots\ldots 1.4.7$

Where

$$F(t) = 8t^5\rho^4 - 4t^4(1-\rho^2)\rho^2\sqrt{\omega/\phi} - 8t^3\rho^2(1-\rho^2)$$

$$+ 4t^2\rho^2(1-\rho^2)\sqrt{\omega/\phi} + 2t\rho^2(\rho^2+3) - \sqrt{\omega/\phi}(1-\rho^4)$$

has a root lying between $\sqrt{1/2}$ and 1.

Now $F(0) = -ve$, $F(\sqrt{1/2}) = 2(4\rho^4+\rho^2) - \sqrt{\omega/\phi}(1-\rho^2)$

$$F(1) = 18\rho^4 - 2\rho^2 - \sqrt{\omega/\phi}(1-\rho^4),\ F(+\infty) = +ve$$

Let $\rho_1^2 > 0$ be the real root of $F(\sqrt{1/2}) = 0$ and $\rho_2^2 > 0$ be the root of $F(1) = 0$, ρ_1^2, ρ_2^2 have been calculated for a wide range of values of ϕ/ω ranging from .001 to 100.00 and are given in table I. If ρ^2 lies between ρ_1^2 and ρ_2^2 then $F(\sqrt{1/2})$ and $F(1)$ have opposite signs and there is a real +ve root between $\sqrt{1/2}$ and 1.

If $\rho_1^2 = \rho_2^2 = \rho'^2$ say

Eliminating ρ^2 from $F(1) = 0$ and $F(\sqrt{1/2}) = 0$ we get

$$125.5384\,\phi/\omega - 49.3014\sqrt{\phi/\omega} - 1.8580 = 0.$$

Solving this equation for $\sqrt{\phi/\omega}$ we get

$$\sqrt{\phi/\omega} = .4274$$

or $\phi/\omega = .18267076$

and $\rho'^2 = 0.3918$

thus if $\phi/\omega = .18267076$

and $\rho'^2 = 0.3918$

we shall adopt either complete replacement or no replacement and variance of the estimate is the same as that of a simple random sample mean of size n.

Real positive roots of equation 1.4.7 for a wide range of ϕ/ω and ρ were obtained with the help of 1620 IBM and results are given in table II.2. Equation 1.4.7 seems to have only one +ve root for all values of ρ^2 and ϕ/ω except when ρ^2 is in the neighbourhood of ρ_1^2 or ρ_2^2. In the latter case there are three +ve values of t, out of which only one root leads to a feasible solution for m. From this table it is clear that t decreases as ρ increases for any value of ϕ/ω. Perhaps this is true for all values of ρ. In that case equation 1.4.7 has a meaningful root i.e. between $\sqrt{1/2}$ and 1 if and only if ρ^2 lies between ρ_1^2 and ρ_2^2. It may also be seen that for $\phi/\omega \geq 1$ equation 1.4.7 does not have a meaningful solution if $\rho \leq .5$ and for $\phi/\omega < 1$ it does not have meaningful solution if $\rho^2 \geq .5$.

CHAPTER II

Regression Estimate in Three-Phase Sampling.

2.1. INTRODUCTION.

With the increase in the cost of enumerating a sampling unit, the number of units which can be enumerated for a given cost may not be large enough to provide estimates with reasonable precision. Especially this will be the case if the standard deviation of the character is high. Regression or ratio estimates are utilised where information on some other auxiliary character is available for the units in the sample together with the total value of the auxiliary character in the population. In case total value of the auxiliary character is not available and it is costly to obtain it, Neyman (1938) has shown that by spending a part of the resources on collecting information for the other correlated character it is possible to increase the precision of the estimate of the population value of the character under study. Information on the correlated character has been used in either of the two ways for this purpose. One way is to utilise information on auxiliary character for stratifying the population and then a sub-sample has been taken within each stratum for measuring the character under study. Alternatively, ancillary information is utilised in developing regression/ratio-estimates. He has given the estimation procedure, the variance of the estimate and the estimate of the variance of the estimate of population mean using stratification or regression/ratio method. Later several other authors such as Bose, C. (1943), (1951), Seal, K.C.(1951), (1953), Sukhatme, B.V. and Kaushal, R.S. (1959), Tikkiwal, B.D. (1960), Goswami, J.N. (1961) and Ray, S.K. (1964), have done a good amount of work involving two-phase sampling, but practically no work has been done in three-phase sampling where use has been made of stratification as well as regression method.

In the present study the estimation procedure and optimum design in three-phase sampling has been obtained. For three-phase sampling it can be easily seen that the double sampling estimate given by Neyman(1938) is a particular case of the estimate suggested in the present investigation.

2.2. Three Phase Sampling: Under the sampling plan proposed here, the sample is drawn in three phases. In the first phase a preliminary sample is drawn to collect information on some character 'X' correlated with the character under study. The population is stratified on the basis of this character and the number of sample units falling within each stratum enumerated. In the second phase, a sample is drawn from the units falling within each stratum to collect information on some other correlated character 'Y'. In the third phase, from each stratum a sub-sample of units on which the character Y is enumerated is drawn for measurement of Z-character under study.

2.3. Notation and Estimation Procedure:

L = number of strata in the population formed on the basis of character X.

N = number of units in the population.

N_h = number of units in the hth stratum in the population.

n' = number of units in the preliminary sample drawn in the first phase.

$W_h = \frac{N_h}{N}$

$Y_{h\alpha}$ = value of the character Y for th unit in the hth stratum in the population.

$Z_{h\alpha}$ = value of character Z for th unit in the hth stratum in the population.

$\bar{Y}_h = \frac{\sum_{\alpha=1}^{N_h} Y_{h\alpha}}{N_h}$, the population mean of the character Y in the hth stratum.

$\bar{Z}_h = \frac{\sum_{\alpha=1}^{N_h} Z_{h\alpha}}{N_h}$, the population mean of the character Z in the hth stratum.

$\bar{Y} = \sum_{h=1}^{L} W_h \bar{Y}_h$, the population mean for character Y.

$\bar{Z} = \sum_{h=1}^{L} W_h \bar{Z}_h$, the population mean for character Z.

Out of N we draw a simple random sample without replacement of size n' and note the number of observations falling in each stratum.

Let n'_h $(h = 1, 2, \ldots\ldots, L)$ denote the number of units falling in the hth stratum so that

$$\sum_{h=1}^{L} n'_h = n'$$

From the hth stratum i.e. out of n'_h $(h=1,2,..L)$ draw a simple random sample of size n_h $(h = 1,2,.....L)$ without replacement.

Where $\sum_{h=1}^{L} n_h = n$, $\quad n < n'$

and on these units measure some other character, Y which is highly correlated with character 'Z' under study. Now, out of n_h $(h=1,2,.....L)$ we take a sub-sample of size m_h $(h=1,2,..L)$ with simple random sample without replacement and we measure character 'Z' under study.

Where $\sum_{h=1}^{L} m_h = m$, $\quad m < n$

We define the estimate of the population mean as

$$\bar{Z} = \sum_{h=1}^{L} w_h \left\{ \bar{z}_h + b_h (\bar{y}'_h - \bar{y}_h) \right\} \qquad ....2.2.1$$

Where,

$$b_h = \frac{\sum_{\alpha=1}^{m_h} (y_{h\alpha} - \bar{y}_h)(z_{h\alpha} - \bar{z}_h)}{\sum_{\alpha=1}^{m_h} (y_{h\alpha} - \bar{y}_h)^2}$$

$$w_h = \frac{n'_h}{n'}$$

$$\bar{z}_h = \frac{1}{m_h} \sum_{\alpha=1}^{m_h} z_{h\alpha}$$

$$\bar{y}'_h = \frac{1}{n_h} \sum_{\alpha=1}^{n_h} y_{h\alpha}$$

$$\bar{y}_h = \frac{1}{m_h} \sum_{\alpha=1}^{m_h} y_{h\alpha}$$

Let us assume the model

$$Z_h = \bar{Z}_h + B_h \left(Y_h - \bar{Y}_h\right) + \epsilon_{h\alpha} \qquad \ldots.2.2.2$$

Where,

$$E\left(\epsilon_{h\alpha}\right) = 0$$

$$V\left(\epsilon_{h\alpha}\right) = S_{he}^2 = S_{hz}^2 \left(1-\rho_h^2\right)$$

Where ρ_h is the correlation coefficient between Y and Z in the hth stratum.

Now $E(\hat{\bar{Z}}) = \bar{Z}$

Thus $\hat{\bar{Z}}$ is an unbiased estimate of population mean.

Let us assume that $y_{h\alpha}$ is normally distributed, then making use of results obtained by Cochran(Sampling Tecnniques) and Sukhatme, P.V. (1953) we get

$$V(\hat{\bar{Z}}) = \sum_{h=1}^{L} \left\{ S_{zh}^2 \left(1-\rho_h^2\right) \left[\frac{1}{m_h} + \left(\frac{1}{m_h} - \frac{1}{n_h}\right) \frac{1}{m_h-3} \right] + \frac{\rho_h^2 S_{Zh}^2}{n_h} \right\} W_h^2$$

$$+ \frac{N-n'}{N-1} \frac{1}{n'} \sum_{h=1}^{L} \left\{ S_{Zh}^2 \left(1-\rho_h^2\right) \left[\frac{1}{m_h} + \left(\frac{1}{m_h} - \frac{1}{n_h}\right) \frac{1}{m_h-3} \right] + \frac{\rho_h^2 S_{Zh}^2}{m_h} \right\} \left\{ W_h \left(1-W_h\right) \right\}$$

$$+ \frac{N-n'}{N-1} \frac{1}{n'} \sum_{h=1}^{L} \bar{Z}_h^2 W_h \left(1-W_h\right)$$

$$- \frac{N-n'}{N-1} \frac{1}{n'} \sum_{h}^{L} \sum_{\substack{h' \\ h \neq h'}}^{L} \bar{Z}_h \bar{Z}_{h'} W_h W_{h'} \qquad \ldots. 2.2.3$$

Now let us assume that

$$\frac{N}{N-1} \simeq 1$$

and $\dfrac{1}{m_h\,(m_h-3)}$ is neglected.

Then from (2.2.3) we have

$$V(\hat{\bar{Z}}) = \sum_{h=1}^{L} \frac{S_{Zh}^{2}\,(1-\rho_h^{2})\,W_h^{2}}{m_h} + \sum_{h=1}^{L} \frac{\rho_h^{2}\,S_{Zh}^{2}\,W_h^{2}}{n_h}$$

$$+ \left(\frac{1}{n'} - \frac{1}{N-1} \right) \sum_{h=1}^{L} W_h\,(\bar{Z}_h - \bar{Z})^{2} \qquad \ldots 2.2.4$$

Where $\bar{Z} = \sum_{h=1}^{L} W_h\,\bar{Z}_h$

<u>Case (i)</u> $m_h = m/L, \quad n_h = n/L$

Then from (2.2.4) we have

$$V(\hat{\bar{Z}}) = \frac{L}{m} \sum_{h=1}^{L} S_{Zh}^{2}(1-\rho_h^{2})\,W_h^{2} + \frac{L}{n} \sum_{h=1}^{L} \rho_h^{2}\,S_{Zh}^{2}\,W_h^{2}$$

$$+ \left(\frac{1}{n'} - \frac{1}{N-1} \right) \sum_{h=1}^{L} W_h\,(\bar{Z}_h - \bar{Z})^{2}$$

$$= \frac{V_1}{m} + \frac{V_2}{n} + \frac{V_3}{n'} - \frac{1}{N-1} \sum_{h=1}^{L} W_h\,(\bar{Z}_h - \bar{Z})^{2} \qquad \ldots 2.2.5$$

Where,

$$V_1 = \sum_{h=1}^{L} S_{Zh}^{2}\,(1-\rho_h^{2})\,W_h^{2}\,L$$

$$V_2 = \sum_{h=1}^{L} \bar{\rho}_h^2 \, S_{zh}^2 \, W_h^2 \, L$$

$$V_3 = \sum_{h=1}^{L} W_h (\bar{Z}_h - \bar{Z})^2$$

Let the cost function be

$$C = C_1 m + C_2 n + C_3 n' \qquad \ldots.2.2.6$$

Let

$$\phi_1 = \left(\frac{V_1}{m} + \frac{V_2}{n} + \frac{V_3}{n'} \right) (C_1 m + C_2 n + C_3 n') \qquad \ldots.2.2.7$$

By Schwartz inequality, the optimum value of variance of the estimate of population mean is given by

$$V_{opt}(\hat{\bar{Z}}) = \frac{(\sqrt{V_1 C_1} + \sqrt{V_2 C_2} + \sqrt{V_3 C_3})^2}{C} - \frac{1}{N-1} \sum_{h=1}^{L} W_h (\bar{Z}_h - \bar{Z})^2 \qquad \ldots.2.2.8$$

and optimum values of m,n,n' are given by

$$n' = \frac{C\sqrt{V_3}}{\sqrt{C_3}\,(\sqrt{V_1 C_1} + \sqrt{V_2 C_2} + \sqrt{V_3 C_3})}$$

$$n = \frac{C\sqrt{V_2}}{\sqrt{C_2}\,(\sqrt{V_1 C_1} + \sqrt{V_2 C_2} + \sqrt{V_3 C_3})}$$

$$m = \frac{C\sqrt{V_1}}{\sqrt{C_1}\,(\sqrt{V_1 C_1} + \sqrt{V_2 C_2} + \sqrt{V_3 C_3})}$$

Case (ii) when m_h, n_h are not same for all strata.

From (2.2.5) we have

$$V(\hat{\bar{Z}}) = \sum_{h=1}^{L} \frac{S_{Zh}^2 (1-\rho_h^2) W_h^2}{m_h} + \sum_{h=1}^{L} \frac{\rho_h^2 S_{Zh}^2 W_h^2}{n_h}$$

$$+ \left(\frac{1}{n'} - \frac{1}{N-1} \right) \sum_{h=1}^{L} W_h (\bar{Z}_h - \bar{Z})^2 \qquad ....2.2.9$$

$$C = C_1 \sum_{h=1}^{L} m_h + C_2 \sum_{h=1}^{L} n_h + C_3 n' \qquad ....2.2.10$$

Minimising $V(\hat{\bar{Z}})$ with respect to m_h, n_h, n' for a fixed cost we get

$$n' = \frac{C \sqrt{\sum_{h=1}^{L} W_h (\bar{Z}_h - \bar{Z})^2}}{A \sqrt{C_3}}$$

$$n_h = \frac{C \rho_h S_{Zh} W_h}{A \sqrt{C_2}}$$

$$m_h = \frac{C S_{Zh} W_h \sqrt{1-\rho_h^2}}{A \sqrt{C_1}}$$

$$V_{opt} (\hat{\bar{Z}}) = \frac{A^2}{C} - \frac{1}{N-1} \sum_{h=1}^{L} W_h (\bar{Z}_h - \bar{Z})^2$$

Where $A = \sqrt{C_1} \sum_{h=1}^{L} S_{Zh} \sqrt{1-\rho_h^2}\, W_h + \sqrt{C_2} \sum_{h=1}^{L} \rho_h S_{Zh} W_h$

$$+ \sqrt{C_3} \sqrt{\sum_{h=1}^{L} W_h (\bar{Z}_h - \bar{Z})^2}$$

2.3 <u>Three Phase Sampling with varying probability at 3rd phase:</u>

When the sample at 3rd phase is drawn with varying probability with replacement.

In the first phase we draw a preliminary sample of size n' and stratify the population on the basis of character X which is correlated with character Z under study. Let n'_h $(h = 1, 2, \ldots L)$ be the number of units falling in the hth stratum. In the second phase out of n'_h $(h = 1, 2, \ldots L)$ a sample of size n_h is selected with simple random sample without replacement and character Y correlated with character Z is measured. Let the observations recorded be as

$y_{h1}, y_{h2}, \ldots\ldots y_{hn_h}$. Then define

$$p_{h\alpha}(n_h) = \frac{y_{h\alpha}}{\sum_{\alpha=1}^{n_h} y_h}, \qquad \alpha = 1, 2, \ldots n_h \quad h = 1, 2, \ldots L \qquad \ldots.2.3.1$$

and from these n_h units select m_h units with the probabilities of selection $p_h(n_h)$, $\alpha = 1, 2, \ldots n_h$ and with replacement. Measure the character Z on these m_h units. Let the observations be denoted by

$z_{h1}, z_{h2}, \ldots\ldots z_{hm_h}$.

Define a new variate

$$z'_{h\alpha} = \frac{z_{h\alpha}}{n_h p_h(n_h)}, \qquad \alpha = 1, 2, \ldots n_h \qquad \ldots.2.3.2$$

Then an estimate for population mean $\bar{Z}$ is given by

$$\hat{\bar{Z}} = \sum_{h=1}^{L} w_h \bar{z}'_h \qquad \ldots.2.3.3$$

Where,

$$\bar{\bar{z}}'_h = \frac{1}{m_h} \sum_{\alpha=1}^{m_h} z'_{h\alpha}$$

$$w_h = \frac{n'_h}{n'}$$

$$\text{Now} \quad E(\hat{\bar{\bar{Z}}}) = \sum_{h=1}^{L} W_h \bar{\bar{Z}}_h \,, \quad W_h = \frac{N_h}{N}$$

$$= \bar{\bar{Z}}$$

where N_h is the number of units in the hth stratum and $N = \sum_{h=1}^{L} N_h$

Thus $\hat{\bar{\bar{Z}}}$ is an unbiased estimate of population mean.

$$V(\hat{\bar{\bar{Z}}}) = \sum_{h=1}^{L} \frac{N_h}{N_h - 1} \left\{ W_h^2 + \frac{N-n'}{N-1} \frac{W_h(1-W_h)}{n'} \right\} x$$

$$\left\{ \frac{n_h - 1}{m_h \; n_h} \sigma^2_{hz'} + \left(\frac{1}{n_h} - \frac{1}{N_h} \right) \sigma^2_{hz} \right\}$$

$$+ \frac{N-n'}{N-1} \sum_{h=1}^{L} W_h (\bar{\bar{Z}}_h - \bar{\bar{Z}})^2 / n' \qquad \ldots 2.3.3$$

Where,

$$\bar{\bar{Z}} = \sum_{h=1}^{L} W_h \bar{\bar{Z}}_h$$

$$\sigma^2_{hz'} = \sum_{i=1}^{N_h} P_{hi} (Z'_{hi} - \bar{\bar{Z}}_{h..})^2 \,, \quad \bar{Z}_{h..} = \sum_{i=1}^{N_h} P_{hi} Z'_{hi}$$

$$\sigma^2_{hz} = \sum_{i=1}^{N_h} \frac{(Z_{hi} - \bar{\bar{Z}}_{h..})^2}{N_h}$$

$$P_{hi} = \frac{Y_{hi}}{\sum_{i=1}^{N_h} Y_{hi}} \,, \quad Z'_{hi} = \frac{Z_{hi}}{N_h P_{hi}}$$

Let us assume that $\frac{N}{N-1} \simeq 1$

and $\frac{1}{m_h^2}$ is neglected.

Then after simplification (2.3.3.) can be written as

$$V(\hat{\bar{\bar{Z}}}) = \sum_{h=1}^{L} \frac{N_h}{N_h-1} \frac{W_h^2 \sigma_{hz'}^2}{m_h} + \sum_{h=1}^{L} \frac{N_h}{N_h-1} \frac{W_h^2 \sigma_{hz}^2}{n_h}$$

$$+ \frac{1}{n'} \sum_{h=1}^{L} W_h (\bar{\bar{Z}}_h - \bar{\bar{Z}})^2 - \sum_{h=1}^{L} \frac{N_h}{N_h-1} \frac{W_h^2 \sigma_{hz}^2}{N_h}$$

$$- \frac{1}{N-1} \sum_{h=1}^{L} W_h (\bar{\bar{Z}}_h - \bar{\bar{Z}})^2 \qquad ....2.3.4$$

Let the cost function be given by

$$C = C_1 \sum_{h=1}^{L} m_h + C_2 \sum_{h=1}^{L} n_h + C_3 n' \qquad ....2.3.5$$

Minimising $V(\bar{\bar{Z}})$ with respect to m_h, n_h, n' for a fixed cost we get

$$n' = \frac{C \sqrt{\sum_{h=1}^{L} W_h (\bar{\bar{Z}}_h - \bar{\bar{Z}})^2}}{B \sqrt{C_3}}$$

$$n_h = \frac{C\, W_h \sigma_{hz}}{B \sqrt{C_2}} \sqrt{\frac{N_h}{N_h-1}} \qquad ....2.3.6$$

$$m_h = \frac{C\, W_h \sigma_{hz'}}{B \sqrt{C_1}} \sqrt{\frac{N_h}{N_h-1}}$$

$$V_{opt}(\hat{\bar{Z}}) = \frac{B^2}{C} - \sum_{h=1}^{L} \frac{N_h}{N_h-1} \frac{W_h^2 \sigma_{hz}^2}{N_h} - \frac{1}{N-1} \sum_{h=1}^{L} W_h (\bar{Z}_h - \bar{Z})^2 \quad \ldots 2.3.7$$

Where,

$$B = \sqrt{C_1} \sum_{h=1}^{L} W_h \sigma_{hz'} \sqrt{\frac{N_h}{N_h-1}} + \sqrt{C_2} \sum_{h=1}^{L} W_h \sigma_{hz} \sqrt{\frac{N_h}{N_h-1}}$$

$$+ \sqrt{C_3} \sum_{h=1}^{L} W_h (\bar{Z}_h - \bar{Z})^2$$

Tables have been prepared to show the efficiency of estimates developed in this chapter as compared to that developed by Neyman (1938) for two phase sampling. For this purpose data was collected from the Institute of Agricultural Research Statistics (I.C.A.R.). The data consists of a list of 124 villages for which character X (area under apple), character Y (no. of orchards) and character Z (apple bearing trees) are available. It was collected during a recent fruit survey conducted by the Institute of Agricultural Research Statistics. These 124 villages are being considered to form the population under study. The whole population was divided into three strata on the basis of character X. Table 3 shows optimum values of n', n_h (h = 1,2, 3), m_h (h = 1,2,3), variance of the estimate with equal probability at all phases (given by equation 2.2.1) and percentage gain in efficiency as compared to the estimate mentioned above for a wide range of 'cost' constants and keeping the total cost to be incurred on the survey same for both the cases. It may be seen from table 3 that percentage gain in efficiency varies from 54% to 63% depending upon various values of cost constants. It may also be seen that for fixed value of C_2 and C_3 the percentage gain in efficiency increases as C_1 increases. Further it may also be mentioned (as is evident from table 3) that n', $n_h(h=1,2,3)$, $m_h(h=1,2,3)$ all decrease as C_1/C_2 increases for a fixed value of C_2/C_3. Table 4 shows, the optimum value of n', $n_h(h=1,2,3)$, $m_h(h=1,2,3)$, optimum value of variance of the estimate with varying probability at the third phase (given by equation 2.3.3) and percentage gain in efficiency as compared to the estimate in two phase sampling. It may be seen that percentage gain in efficiency varies from 57% to 67% depending upon various values of cost constants.

It may also be seen that for fixed value of C_2/C_3 percentage gain in efficiency for both equal and varying probability increases as C_1/C_2 increases.

In the end, it may also be mentioned that as expected percentage gain in efficiency is somewhat more when the sample at the third phase is drawn with varying probability than that when simple random sampling is adopted at the third phase.

2.4 Multiple Regression Estimate.

Now we shall utilize the information on character X for developing regression estimate instead of utilizing it for stratifying the population. The sampling plan is the same as given in (2.2) except that in the second phase we draw a sample of size n out of unstratified sample of size n' and in the third phase we draw a sample of size m out of unstratified sample of size n, simple random sampling without replacement being adopted at each phase.

Let us define the estimate of population mean as given below.

$$\hat{\bar{Z}} = \bar{z} + b_{zy}\,(\bar{y}' - \bar{y}) + b_{zx}\,(\bar{x}' - \bar{x}) \qquad \ldots.2.4.1$$

Where,

$$\bar{z} = \frac{1}{m}\sum_{i=1}^{m} z_i$$

$$\bar{y}' = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$\bar{x}' = \frac{1}{n'}\sum_{i=1}^{n} x_i$$

Variance of $\hat{\bar{Z}}$ is minimum when

$$b_{zx} = \frac{\mathrm{Cov}\left\{\bar{z},(\bar{y}'-\bar{y})\right\}\mathrm{Cov}\left\{(\bar{y}'-\bar{y}),(\bar{x}'-\bar{x})\right\} - \mathrm{Cov}\left\{\bar{z},(\bar{x}'-\bar{x})\right\}V(\bar{y}'-\bar{y})}{V(\bar{x}'-\bar{x})\,V(\bar{y}'-\bar{y}) - \left\{\mathrm{Cov}\left[(\bar{y}'-\bar{y}),(\bar{x}'-\bar{x})\right]\right\}^2} \qquad \ldots.2.4.2$$

$$b_{zy} = \frac{\text{cov}\left\{\bar{z}, (\bar{x}'-\bar{x})\right\} \text{cov}\left\{(\bar{y}'-\bar{y}), (\bar{x}'-\bar{x})\right\} - \text{cov}\left\{\bar{z}, (\bar{y}'-\bar{y})\right\} V(\bar{x}'-\bar{x})}{V(\bar{x}'-\bar{x}) V(\bar{y}'-\bar{y}) - \left\{\text{cov}\left[(\bar{y}'-\bar{y}), (\bar{x}'-\bar{x})\right]\right\}} \quad \ldots\ldots 2.4.3$$

Now if the population values are not known we put the corresponding values estimated from the sample.

$$V(\bar{z}) = \left(\frac{1}{m} - \frac{1}{N}\right) s_Z^2$$

$$V(\bar{y}'-\bar{y}) = \left(\frac{1}{m} - \frac{1}{n}\right) s_Y^2$$

$$V(\bar{x}'-\bar{x}) = \left(\frac{1}{m} - \frac{1}{n'}\right) s_X^2 \quad \ldots\ldots\ldots\ldots 2.4.4$$

$$\text{cov}\left[\bar{z}, (\bar{y}'-\bar{y})\right] = -\left(\frac{1}{m} - \frac{1}{n}\right) s_{YZ}$$

$$\text{cov}\left[\bar{z}, (\bar{x}'-\bar{x})\right] = -\left(\frac{1}{m} - \frac{1}{n'}\right) s_{ZX}$$

$$\text{cov}\left[(\bar{y}'-\bar{y}), (\bar{x}'-\bar{x})\right] = \left(\frac{1}{m} - \frac{1}{n}\right) s_{YX}$$

putting the values from (2.4.4) in (2.4.2) and (2.4.3) we have

$$b_{zx} = \frac{\left(\frac{1}{m} - \frac{1}{n'}\right) s_{ZX} s_Y^2 - \left(\frac{1}{m} - \frac{1}{n}\right) s_{YZ} s_{YX}}{\left(\frac{1}{m} - \frac{1}{n'}\right) s_X^2 s_Y^2 - \left(\frac{1}{m} - \frac{1}{n}\right) s_{YX}^2} \quad \ldots\ldots 2.4.5$$

$$b_{zy} = \frac{\left(\frac{1}{m} - \frac{1}{n'}\right) s_{YZ} s_X^2 - \left(\frac{1}{m} - \frac{1}{n}\right) s_{ZX} s_{YX}}{\left(\frac{1}{m} - \frac{1}{n'}\right) s_X^2 s_Y^2 - \left(\frac{1}{m} - \frac{1}{n}\right) s_{YX}^2} \quad \ldots\ldots 2.4.6$$

Let $m = fn$

$n = gn'$ 2.4.7

putting these values of m & n in (2.4.6) and (2.4.6) we have

$$b_{zx} = \frac{S_Z\left\{(1-fg)\,\rho_{zx} - (1-f)\,\rho_{zy}\,\rho_{xy}\right\}\left\{1 - \frac{fg-f\rho^2_{xy}}{1-\rho^2_{xy}}\right\}^{-1}}{S_X\,(1-\rho^2_{xy})}$$

where,

$$\rho_{zx} = \frac{S_{ZX}}{S_Z\,S_X}$$

Similarly we define ρ_{xy} & ρ_{zy}.

or

$$b_{zx} = \frac{S_Z\left\{(1-fg)\,\rho_{zx} - (1-f)\,\rho_{zy}\,\rho_{xy}\right\}}{S_X\,(1-\rho^2_{xy})}\left\{1 + \frac{fg-f\rho^2_{xy}}{1-\rho^2_{xy}} + \frac{f^2g^2 - 2f^2g\,\rho^2_{xy} + f^2\rho^4_{xy}}{(1-\rho^2_{xy})^2} - \frac{f^3\rho^6_{xy}}{(1-\rho^2_{xy})^3} + \frac{f^4\rho^8_{xy}}{(1-\rho^2_{xy})^4}\right\}$$

neglecting 5th and higher powers of f,g we have

$$b_{zx} = \frac{S_Z}{S_X(1-\rho^2_{xy})}\left\{\rho_{zx.y} + f\,\frac{\rho_{xy}\,(\rho_{zy.x})}{1-\rho^2_{xy}} - fg\,\frac{\rho_{xy}\,(\rho_{zy.x})}{1-\rho^2_{xy}} + - f^2\,\frac{\rho^3_{xy}\,\rho_{zy.x}}{(1-\rho^2_{xy})^2} + f^2g\,\frac{\rho_{xy}\,\rho_{xy.x}(1+\rho_{xy})}{(1-\rho^2_{xy})^2} + \left[\frac{\rho_{zy}\,\rho^5_{xy}}{(1-\rho^2_{xy})^2} - \frac{\rho^6_{xy}\,\rho_{zx.y}}{(1-\rho^2_{xy})^3}\right]\right.$$

$$+f^3 g \frac{-2 \rho_{zy} \rho_{xy}^3 - \rho_{xy}^4 \rho_{zx}}{(1-\rho_{xy}^2)^2} - f^2 g^2 \frac{\rho_{xy} \rho_{zy.x}}{(1-\rho_{xy}^2)^2}$$

$$+ \left[\frac{\rho_{xy}^8 \rho_{zx.y}}{(1-\rho_{xy}^2)^4} - \frac{\rho_{xy}^7 \rho_{zy}}{(1-\rho_{xy}^2)^3} \right] f^4 \Bigg\} \qquad \ldots.2.4.8$$

Similarly

$$b_{zy} = \frac{S_Z \rho_{zy.x}}{S_Y (1-\rho_{xy}^2)} \Bigg\{ 1 - \frac{\rho_{xy}^2}{1-\rho_{xy}^2} f + \frac{\rho_{xy}^2}{1-\rho_{xy}^2} fg + \frac{\rho_{xy}^4}{(1-\rho_{xy}^2)^2} f^2$$

$$- \frac{\rho_{xy}^2 (1+\rho_{xy}^2)}{(1-\rho_{xy}^2)^2} f^2 g + \frac{\rho_{xy}^2}{(1-\rho_{xy}^2)^2} f^2 g^2$$

$$- \frac{\rho_{xy}^6}{(1-\rho_{xy}^2)^3} f^3 - \frac{\rho_{xy}^4}{(1-\rho_{xy}^2)^2} f^3 g + \frac{\rho_{xy}^8}{(1-\rho_{xy}^2)^4} f^4 \Bigg\} \qquad \ldots..2.4.9$$

where $\rho_{zy.x} = \rho_{zy} - \rho_{zx} \rho_{yx}$, similarly we define $\rho_{zx.y}$

Now

$$V(\hat{Z}) = \frac{1}{nf} \Big\{ S_Z^2 + b_{zy}^2 (1-f) S_Y^2 + b_{zx}^2 (1-fg) S_X^2 + 2 b_{zy} (1-f) S_{YZ}$$

Contd....

$$-2b_{zx}(1-fg)S_{ZX} + 2b_{zx}\, b_{zy}(1-f)S_{YX} \Big\} - S_Z^2/N \qquad \ldots\ldots 2.4.10$$

Let the cost function be

$$C = C_1 m + C_2 n + C_3 n' \qquad \ldots\ldots 2.4.11$$

$$= \frac{n}{g}(C_1 fg + C_2 g + C_3)$$

Putting the value of n from (2.4.11) in (2.4.10) we have

$$V(\hat{\bar{Z}}) = \frac{1}{Cfg}(C_1 fg + C_2 g + C_3)\Big\{S_Z^2 + b_{zy}^2(1-f)S_Y^2 + b_{zx}^2(1-fg)S_X^2$$

$$-2b_{zy}(1-f)S_{YZ} - 2b_{zx}(1-fg)S_{ZX}$$

$$+2b_{zx}b_{zy}(1-f)S_{YX}\Big\} - S_Z^2/N$$

$$\ldots\ldots 2.4.12$$

Minimizing $V(\hat{\bar{Z}})$ from (2.4.12) with respect to f and g we get the optimum values of f and g. Then from (2.4.10), (2.4.7) and (2.4.11) we get optimum values of n, n', m and variance of the estimate.

APPENDIX I

The variance of the regression estimate of the population mean of the character under study is underestimated when the regression coefficient is assumed to be constant when in fact it is subject to sampling fluctuations. For single stage sampling design this under estimation was studied by Cochran (sampling techniques). In this note it is planned to study when a two-stage sampling design is adopted.

Let

Y_{ij} = value of the character under study for the jth secondary unit in the ith primary unit.

X_{ij} = value of the auxiliary character for the jth secondary unit in the ith primary unit.

N = total no. of primary units in the population.

M = total no. of secondary units in each primary unit.

n = no. of primary stage units selected.

m = no. of secondary units selected from each primary stage unit selected.

$$\bar{Y}_{NM} = \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} Y_{ij}$$

$$\bar{X}_{NM} = \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} X_{ij}$$

$$\bar{y}_{i.} = \frac{1}{m} \sum_{j=1}^{m} y_{ij}$$

M

$$\bar{x}_{i.} = \frac{1}{m} \sum_{j=1}^{m} x_{ij}$$

$$\bar{y}_{..} = \frac{1}{n} \sum_{i=1}^{n} \bar{y}_{i.}$$

$$\bar{x}_{..} = \frac{1}{n} \sum_{i=1}^{n} \bar{x}_{i.}$$

We define the regression estimate of $\bar{Y}_{NM}$ as

$$\bar{y}_1 = \bar{y}_{..} + \beta (\bar{X}_{NM} - \bar{x}_{..}) \qquad ......(1)$$

where β is to be so determined that $V(\bar{y}_1)$ is minimum; and this value of β is given by

$$\beta = \frac{\mathrm{cov}(\bar{y}_{..}, \bar{x}_{..})}{V(\bar{x}_{..})}$$

As β is generally not known we replace it by its estimate 'b' given by

$$b = \frac{\widehat{\mathrm{cov}}(\bar{y}_{..}, \bar{x}_{..})}{\hat{V}(\bar{x}_{..})}$$

where ' $\wedge$ ' indicates the usual unbiased estimate of the population parameters.

Now

$$\mathrm{cov}(\bar{y}_{..}, \bar{x}_{..}) = \left(\frac{1}{n} - \frac{1}{N}\right) S_{byx} + \frac{1}{nN}\left(\frac{1}{m} - \frac{1}{M}\right) \sum_{i=1}^{N} S_{iyx}$$

$$V(\bar{x}_{..}) = \left(\frac{1}{n} - \frac{1}{N}\right) S_{bx}^2 + \frac{1}{nN}\left(\frac{1}{m} - \frac{1}{M}\right) \sum_{i=1}^{N} S_{ix}^2$$

where

$$S_{byx} = \frac{1}{N-1} \sum_{i=1}^{N} (\bar{Y}_{i.} - \bar{Y}_{NM})(\bar{X}_{i.} - \bar{X}_{..})$$

$$S_{iyx} = \frac{1}{M-1} \sum_{j=1}^{M} (Y_{ij} - \bar{Y}_{i.})(X_{ij} - \bar{X}_{i.})$$

$$S_{bx}^2 = \frac{1}{N-1} \sum_{i=1}^{N} (\bar{X}_{i.} - \bar{X}_{..})^2$$

$$s_{ix}^2 = \frac{1}{M-1} \sum_{j=1}^{M} (X_{ij} - \bar{X}_{i.})^2$$

If we assume N is large such that terms involving N^{-a} where $a \geq 1$ can be neglected then

$$\hat{cov} (\bar{y}_{...}, \bar{x}_{...}) \doteq \frac{s_{byx}}{n}$$

$$\hat{V}(\bar{y}_{..}) \doteq \frac{s_{by}^2}{n}$$

where

$$s_{byx} = \frac{1}{n-1} \sum_{i=1}^{n} (\bar{y}_{i.} - \bar{y}_{..})(\bar{x}_{i.} - \bar{x}_{..})$$

$$s_{bx}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\bar{x}_{i.} - \bar{x}_{..})^2$$

$$\therefore b = \frac{s_{byx}}{s_{bx}^2}, \text{ neglecting } 1/N \text{ and higher powers of } 1/N \qquad \ldots\ldots(2)$$

Variance of $\bar{y}_1$ in (1), for a given value of $\beta = b$ is given by

$$V(\bar{y}_{1/b}) = V(\bar{y}_{..}) + b^2 V(\bar{x}_{..}) - 2b \, cov(\bar{y}_{..}, \bar{x}_{..}) \qquad \ldots\ldots.(3)$$

where 'b' is given by (2).

Variance of $\bar{y}_1$ when β is replaced by 'b' as given by (2) is given by

$$V_1(\bar{y}_1) = E (V\bar{y}_1/b) + V(E\bar{y}_1/b)$$

The fixing of 'b' does not affect the distribution of $\bar{y}_{..}$ and $\bar{x}_{..}$ in case of large samples specially when (x,y) follow a normal distribution.

Thus the underestimation of the variance of the regression estimate is given by $V(E\bar{y}_1/b)$.

Now $V(E\bar{y}_1/b) = 0$

Thus we may conclude that under-estimation of the variance of the regression estimate of the population mean of the character under-study when the regression coefficient is assumed to be constant when in fact it is subject to sampling fluctuations is negligible when sample size is large specially when 'x', 'y' follow a normal distribution.

****

Table I

Table showing the values of ρ_2^2 and ρ_1^2 for various values of q/ω

q/ω	ρ_2^2	ρ_1^2	q/ω	ρ_2^2	ρ_1^2
.001	.81869412	.83718340	.051	.49115435	.50803120
.002	.76951497	.79146910	.052	.48952736	.50620680
.003	.73772884	.76139960	.053	.48793429	.50441800
.004	.71405352	.73870070	.054	.48637388	.50266360
.005	.69515650	.72038180	.055	.48484490	.50094240
.006	.67942993	.70499170	.056	.48334619	.49925305
.007	.66596753	.69170770	.057	.48187666	.49759461
.008	.65420570	.68001530	.058	.48043524	.49596596
.009	.64376911	.66957100	.059	.47902098	.49436606
.010	.63439479	.66013200	.060	.47763296	.49279394
.011	.62589118	.65152140	.061	.47627025	.49124881
.012	.61811428	.64360540	.062	.47493204	.48972969
.013	.61095311	.63628060	.063	.47361749	.48823576
.014	.60432028	.62946500	.064	.47232586	.48676628
.015	.59814573	.62309300	.065	.47105643	.48532042
.016	.59237238	.61711090	.066	.46980846	.48389759
.017	.58695322	.61147420	.067	.46858133	.48249697
.018	.58184898	.60614550	.068	.46737439	.48111801
.019	.57702651	.60109350	.069	.46618704	.47976000
.020	.57245763	.59629120	.070	.46501865	.47842243
.021	.56811816	.59171550	.071	.46386875	.47710464
.022	.56398718	.58734630	.072	.46273674	.47580612
.023	.56004647	.58316620	.073	.46162214	.47452631
.024	.55628013	.57915980	.074	.46052444	.47326475
.025	.55267412	.57531340	.075	.45944322	.47202093
.026	.54921598	.57161530	.076	.45837797	.47079437
.027	.54589472	.56805440	.077	.45732828	.46958459
.028	.54270042	.56462130	.078	.45629373	.46839125
.029	.53962426	.56130740	.079	.45527395	.46721386
.030	.53665827	.55810490	.080	.45426853	.46605200
.031	.53379526	.55500670	.081	.45327710	.46490534
.032	.53102870	.55200630	.082	.45229932	.46377349
.033	.52835266	.54909810	.083	.45133481	.46265607
.034	.52576174	.54627670	.084	.45038329	.46155277
.035	.52325100	.54353710	.085	.44944440	.46046323
.036	.52081590	.54087500	.086	.44851787	.45938713
.037	.51845226	.53828610	.087	.44760337	.45832414
.038	.51615630	.53576680	.088	.44670063	.45727403
.039	.51392446	.53331340	.089	.44580938	.45623640
.040	.51175344	.53092270	.090	.44492934	.45521108
.041	.50964028	.52859180	.091	.44406028	.45419775
.042	.50758207	.52631770	.092	.44320187	.45319609
.043	.50557625	.52409790	.093	.44235397	.45220594
.044	.50362040	.52193000	.094	.44151628	.45122701
.045	.50171222	.51981160	.095	.44068862	.45025910
.046	.49984959	.51774060	.096	.43987076	.44930189
.047	.49803053	.51571510	.097	.43906245	.44835522
.048	.49625317	.51373310	.098	.43826355	.44741884
.049	.49451577	.51179290	.099	.43747379	.44649259
.050	.49281669	.50989270	.100	.43669305	.44557627

Contd.....

ϕ/ω	ρ_2^2	ρ_1^2	ϕ/ω	ρ_2^2	ρ_1^2
.110	.42934109	.43691611	.510	.32510604	.30748922
.120	.42271306	.42905943	.520	.32396279	.30599522
.130	.41668778	.42187614	.530	.32284558	.30453362
.140	.41117162	.41526502	.540	.32175339	.30310316
.150	.40609078	.40914582	.550	.32068520	.30170258
.160	.40138612	.40345393	.560	.31964009	.30033081
.170	.39700956	.39813648	.570	.31861717	.29898673
.180	.39292146	.39314977	.580	.31761559	.29766932
.190	.38908882	.38845723	.590	.31663457	.29637773
.200	.38548387	.38402785	.600	.31567336	.29511089
.210	.38208304	.37983532	.610	.31473121	.39386802
.220	.37886614	.37585697	.620	.31380750	.29264829
.230	.37581575	.37207327	.630	.31290155	.29145095
.240	.37291685	.36846710	.640	.31201278	.29027517
.250	.37015621	.36502952	.650	.31114059	.28912030
.260	.36752221	.36172940	.660	.31028441	.28798572
.270	.35500472	.35857300	.670	.30944378	.28687066
.280	.36259463	.35554403	.680	.30861812	.28577459
.290	.36028386	.35263321	.690	.30780701	.28469694
.300	.35806523	.34983225	.700	.30700996	.28363708
.310	.35593224	.34713366	.710	.30622655	.28259454
.320	.35387905	.34453075	.720	.30545637	.28156874
.330	.35190042	.34201730	.730	.30469901	.28055927
.340	.34999153	.33958778	.740	.30395408	.27956559
.350	.34814806	.33723723	.750	.30322124	.27858734
.360	.34636601	.33496089	.760	.30250012	.27762397
.370	.34464181	.33275456	.770	.30179042	.27667514
.380	.34297209	.33061438	.780	.30109176	.27574046
.390	.34135383	.32853679	.790	.30040389	.27481952
.400	.33978421	.32651836	.800	.29972647	.27391195
.410	.33826062	.32455610	.810	.29905924	.27301741
.420	.33678067	.32264727	.820	.29840192	.27213555
.430	.33534217	.32078909	.830	.29775425	.27126607
.440	.33394303	.31897926	.840	.29711597	.27040862
.450	.33258135	.31721544	.850	.29648685	.26956292
.460	.33125536	.31549550	.860	.29586664	.26872868
.470	.32996337	.31381745	.870	.29525513	.26790561
.480	.32870386	.31217952	.880	.29465209	.26709344
.490	.32747534	.31057989	.890	.29405733	.26629191
.500	.32627631	.30901695	.900	.29347061	.26550077

Contd....

Φ/μ	ρ_2^2	ρ_1^2	Φ/μ	ρ_2^2	ρ_1^2
.010	.29289177	.26471976	46.000	.16075047	.07437583
.020	.29232050	.26334865	47.000	.16034908	.7376721
.030	.29175693	.26313724	48.000	.15995861	.07317521
.040	.29120059	.26043527	49.000	.15957871	.07259011
.050	.29065140	.26108254	50.000	.15920888	.07203310
.060	.29010913	.26025337	51.000	.15884861	.07149174
.070	.28957331	.26002602	52.000	.15849751	.07096915
.080	.28904512	.25951782	53.000	.15815513	.07046380
.090	.28852294	.25881007	54.000	.15782123	.06998815
1.000	.28800716	.25811062	55.000	.15749530	.06952364
2.000	.25503977	.21269536	56.000	.15717709	.06905880
3.000	.23315407	.18372778	57.000	.15686627	.06862413
4.000	.22711145	.17258041	58.000	.15656254	.06820828
5.000	.21907674	.16124234	59.000	.15626463	.06757264
6.000	.21284320	.15215188	60.000	.15597520	.06713199
7.000	.20779793	.14475615	61.000	.15569125	.06670090
8.000	.20358981	.13856191	62.000	.15541328	.06627900
9.000	.19999993	.13325985	63.000	.15514117	.06586597
10.000	.19688364	.12864400	64.000	.15487470	.06546149
11.000	.19414231	.12459070	65.000	.15461367	.06506525
12.000	.19169733	.12093606	66.000	.15435789	.06467697
13.000	.18950113	.11766262	67.000	.15410718	.06429633
14.000	.18751099	.11469126	68.000	.15386135	.06392321
15.000	.18569499	.11197589	69.000	.15362027	.06355722
16.000	.18402713	.10947985	70.000	.15338376	.06319817
17.000	.18243623	.10717366	71.000	.15315169	.06284583
18.000	.18100134	.10503321	72.000	.15292328	.06250000
19.000	.17973238	.10303857	73.000	.15270024	.06216046
20.000	.17849148	.10117306	74.000	.15248061	.06182702
21.000	.17733734	.09943264	75.000	.15226488	.06149949
22.000	.17623394	.09777537	76.000	.15205292	.06117770
23.000	.17520074	.09622101	77.000	.15184463	.06086146
24.000	.17422493	.09475070	78.000	.15163990	.06055062
25.000	.17330031	.09335676	79.000	.15143963	.06024503
26.000	.17242233	.09203290	80.000	.15124069	.05994452
27.000	.17158703	.09077202	81.000	.15104600	.05964895
28.000	.17078097	.08957010	82.000	.15085440	.05935819
29.000	.17003076	.08842216	83.000	.15066605	.05907210
30.000	.16930790	.08732401	84.000	.15048061	.05879054
31.000	.16860753	.08627207	85.000	.15029807	.05851342
32.000	.16794041	.08526300	86.000	.15011836	.05824060
33.000	.16729353	.08429384	87.000	.14994141	.05797198
34.000	.16668344	.08336189	88.000	.14976715	.05770730
35.000	.16609049	.08246473	89.000	.14959550	.05744630
36.000	.16551931	.08160013	90.000	.14942639	.05719003
37.000	.16496225	.08076611	91.000	.14925976	.05693711
38.000	.16443637	.07996680	92.000	.14909555	.05668788
39.000	.16392266	.07918254	93.000	.14893371	.05644216
40.000	.16342547	.07842077	94.000	.14877416	.05620093
41.000	.16294443	.07770105	95.000	.14861686	.05596116
42.000	.16247852	.07699510	96.000	.14846174	.05572570
43.000	.16202688	.07631069	97.000	.14830877	.05549349
44.000	.16158878	.07564670	98.000	.14815780	.05526445
45.000	.16116352	.07500209	99.000	.14800904	.05503851
			100.000	.14786217	.05481561

Table II

Table showing the real +ve roots of equation 1.4.7 for a given range of ϕ/μ & ρ

ϕ/μ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	ρ_2^2	ρ_1^2
0.003	933.85	219.19	92.40	48.01	27.46	16.28	9.50	5.04	0.5870	.7377	.7614
0.005	700.16	169.84	71.62	37.24	21.30	12.63	7.36	3.84	0.5119	.6952	.7204
0.006	639.19	155.07	65.41	34.01	19.47	11.54	6.72	3.47	0.4865	.6794	.7050
0.008	553.60	134.34	56.68	29.49	16.89	10.01	5.82	2.94	0.4472	.6542	.6800
0.010	495.20	120.19	50.74	26.41	15.13	8.97	5.20	0.7806 1.0500 2.5600	0.4174	.6344	.6601
0.055	211.52	51.61	21.97	11.56	6.69	3.97	0.7561 2.1091 1.0527	0.4332	0.2197	.4848	.5009
0.075	181.27	44 .34	18.94	10.01	5.82	3.45	0.6461 1.3927 1.5695	0.3874	0.1917	.4594	.4720
0.150	128.56	31.71	13.72	7.36	4.33	2.54	0.4890	0.2952	0.1394	.4061	.4091
0.180	117.49	29.07	12.64	6.81	4.03	2.36	0.4547	0.2734	0.1279	.3929	.3931
0.190	114.40	28.34	12.34	6.66	3.95	2.30	0.4449	0.2672	0.1247	.3891	.3884
0.600	65.40	16.86	7.71	4.37	2.61	0.4462	0.2719	0.1588	0.0713	.3157	.2951

Φ/μ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	ρ_2^2	ρ_1^2
0.750	58.82	15.36	7.12	4.08	2.50	0.4008	0.2456	0.1429	0.0639	0.3032	0.2786
1.000	51.41	13.68	6.47	3.76	2.32	0.3492	0.2149	0.1244	0.0554	.2880	.2581
4.000	28.22	8.67	4.55	2.82	0.2929 1.1240 1.7071	0.1783	0.1101	0.0631	0.0278	.2271	.1729
10.000	20.45	7.07	3.94	0.3200 0.8669 2.5140	0.1833	0.1133	0.0700	0.0400	0.0176	.1969	.1286
20.000	16.88	6.34	3.66	0.2214 0.9419 2.3634	0.1293	0.0803	0.0496	0.0283	0.0125	.1785	.1012
25.000	16.02	6.16	0.4321 0.6783 3.5850	0.1968 0.9595 2.3254	0.1158	0.0718	0.0443	0.0253	0.0111	.1733	.0934
60.000	13.60	5.65	0.2428 0.8202 3.3799	0.1254 1.0112 2.2123	0.0745	0.0464	0.0287	0.0164	0.0072	.1560	.0671
75.000	13.17	5.56	0.2145 0.8390 3.3419	0.1120 1.0310 2.1908	0.0663	0.0415	0.0256	0.0146	0.0064	.1523	.0615
90.000	12.85	5.49	0.1943 0.8521 3.3133	0.1021 1.0283 2.1747	0.0608	0.0379	0.0234	0.0134	0.0059	.1494	.0519

Table III

Table showing optimum values of n', $n_h (h = 1,2,3)$, $m_h (h = 1,2,3)$, variance of the estimate with equal probability at all phases (given by equation 2.2.1) and percentage gain in efficiency as compared to the regression estimate developed by Neyman (1938) for two phase sampling.

$C/C_3 = 500$

C_2/C_3	C_1/C_3	n'	n_1	m_1	n_2	m_2	n_3	m_3	$V(\bar{\bar{z}})$	% gain in efficiency
2	4	119	9	12*	28	12	29	7	20525	54.88
2	6	107	8	9*	25	9	26	5	37413	56.79
2	8	99	7	7	23	7	24	4	45597	58.13
2	10	93	7	6	21	6	22	3	53343	59.25
2	12	87	6	5	20	5	21	3	60700	60.11
2	14	83	6	4	19	4	20	3	70032	60.83
2	16	79	6	4	18	4	19	2	75058	61.42
3	4	103	6	6	19	6	20	5	43394	55.37
3	6	91	6	6	17	6	18	4	55039	57.58
3	8	84	5	5	16	5	16	3	60000	59.09
3	10	78	5	4	15	4	15	3	73116	60.22
3	12	73	4	3	14	3	15	2	83032	61.10
3	14	70	4	3	13	3	14	2	93040	61.83
3	16	67	4	3	13	3	13	2	100700	62.43
4	4	90	5	6	15	6	18	4	55720	56.60
4	6	81	4	5	13	5	14	3	72220	58.20
4	8	75	4	4	12	4	13	2	87383	59.33
4	10	69	4	3	11	3	12	2	102229	60.95
4	12	65	3	3	11	3	11	2	116361	61.83
4	14	63	3	3	10	3	11	1	130106	62.54
4	16	59	3	3	10	3	10	1	142344	63.13
5	4	83	4	5	12	5	13	3	63759	56.57
5	6	75	3	4	11	4	11	3	80003	58.88
5	8	67	3	3	10	3	10	2	107335	60.41
5	10	63	3	3	9	3	10	1	125924	61.52
5	12	59	3	3	9	3	9	1	143340	62.38
5	14	56	3	3	8	3	9	1	160292	63.08
5	16	53	3	3	8	3	8	1	170870	63.65

* we shall take $m_1 = n_1$

Table IV

Table showing optimum values of n', n_h ($h = 1, 2, 3$), m_h ($h = 1, 2, 3$), variance of the estimate with varying probability at the third phase (given by equation 2.3.3) and percentage gain in efficiency as compared to the regression estimate developed by Neyman (1938) for two - phase sampling.

$C/C_3 = 500$

C_2/C_3	C_1/C_2	n'	n_1	m_1	n_2	m_2	n_3	m_3	$V(\bar{Z})$	% gain in efficiency
2	4	117	25	8	30	8	32	9	21720	65.71
2	6	107	23	6	28	6	29	7	29987	65.37
2	8	100	21	5	26	5	27	6	37531	65.57
2	10	94	20	4	25	4	25	5	44514	65.92
2	12	90	19	4	23	4	24	4	51368	66.30
2	14	86	18	3	22	3	23	4	57873	66.68
2	16	83	18	3	22	3	22	3	64182	67.03
3	4	100	18	6	21	6	22	6	37484	60.55
3	6	91	16	4	19	4	20	5	49376	61.99
3	8	85	15	4	18	3	19	4	60264	63.15
3	10	80	14	3	17	3	18	3	70509	64.09
3	12	76	13	3	16	3	17	3	80299	64.88
3	14	72	13	2	15	2	16	3	89742	65.55
3	16	70	12	2	15	2	15	2	98912	66.13
4	4	89	13	5	16	4	17	5	52715	58.39
4	6	81	12	3	15	3	15	4	68167	60.64
4	8	75	11	3	14	3	14	3	82343	62.24
4	10	71	11	2	13	2	13	3	95703	63.45
4	12	67	10	2	12	2	13	2	108483	64.42
4	14	64	10	2	12	2	12	2	120825	65.21
4	16	61	9	2	11	2	12	2	132819	65.89
5	4	81	11	4	13	4	14	4	67606	57.30
5	6	73	10	3	12	3	13	3	86575	60.01
5	8	68	9	2	11	2	12	2	104004	61.84
5	10	64	9	2	11	2	11	2	120447	63.20
5	12	61	8	2	10	2	10	2	136192	64.26
5	14	58	8	1	10	1	10	2	151406	65.13
5	16	56	8	1	9	1	10	1	166200	65.85

SUMMARY

The following problems have been attempted in the present work.

I. Investigation of optimum design in two-stage sampling when the survey is to be conducted on two successive occasions.

II. To determine the estimation procedure and investigation of an optimum design in three-phase sampling.

As regards problem I, attempt has been made to obtain expression for optimum allocation and sampling fraction for a two-stage sampling design when the survey is to be conducted on two successive occasions and the expenditure to be incurred does not change from one occasion to another. For the sake of simplicity the discussion is confined to a two-stage design where all the sampling units at both the stages were drawn ~~wik~~ with equal probability and without replacement and out of the primary stage units examined at the first occasion a fraction 'p' was retained for the second occasion.

As regards problem II, the concept of three-phase sampling has been introduced where use has been made of stratification as well as regression method. The estimation procedure and optimum design in three-phase sampling has been obtained under two sampling designs.

(i) when sample at each phase is drawn with equal probability.

(ii) when sampling units at third phase are drawn with varying probability.

Tables have been prepared to show the efficiency of

estimates developed here to compare with the regression estimate developed by Neyman (1938) for two-phase sampling, the total cost to be incurred in both the cases being same. From the tables it is clear that percentage gain in efficiency in the case of estimate with equal probability at each phase varies from 54% to 63% depending upon various values of 'cost' constants & the percentage gain in efficiency in the case of estimate with varying probability at third phase varies from 57% to 67% depending upon various values of 'cost' constants.

It has also been shown that the underestimation of the variance of regression estimate of the population mean of the character under study when the regression coefficient is assumed to be constant when in fact it is subject to sampling fluctuations is negligible when sample size is large specially when 'X' (auxiliary character) and 'Y' (character under study) are normally distributed.

R_E_F_E_R_E_N_C_E_S

(1) Bose, C. (1943) Note on the sampling error in the method of double sampling. Sankhya, 6, 329-330.

(2) - (1951) Some further results on errors in double sampling technique; Sankhya, 11, 191-194.

(3) Cochran, W. G. (1953) Sampling Techniques - Joh Willey & Sons, New York.

(4) Eckler, R. A. (1955) "Rotation Sampling", Ann. Math. Stat., 26.

(5) Goswami, J.N. (1961) On some extensions of double sampling unpublished thesis for Diploma, I.C.A.R.

(6) Hansen, M.H., Hurwitz, W.N., and Madow, W.G. (1953) Sample Survey Methods and Theory, Vol. 182. New York, John Willey & Sons, Inc.

(7) Jessen, R.J. (1942) "Statistical investigation of a sample survey for obtaining farm facts", Iowa Agri. Exp. Stat. Res. Bull., 304.

(8) Kathuria, O.P. (1959) "Some Asepects of Successive Sampling in Multistage Designs", (unpublished thesis submitted towards fulfilment of requirements for Diploma, I.C.A.R., New Delhi.)

(9) Narain, R.D. (1953) "On the Recurrence Formula in Sampling on Successive occasions", J. Ind. Soc. Agri. Stat., 5.

(10) Jeyman, J. (1938) "Contribution to the theory of Sampling Human Populations"., J.A.S.A. 1938, Vol. 33, 101-116.

(11) Patterson, H.D. (1950) "Sampling on Successive Occasions with partial replacement of units", J.Roy. Stat. Soc., Series B, 12.

(12) Ray S.K. (1964) "Regression Technique in Multi-stage Design", (unpublished thesis submitted towards fulfilment of requirements for Diploma, I.C.A.R., New Delhi.)

(13) Sawdasia, V.B. (1964) "Asymmetric Rotation Designs in Sampling on Successive Occasions", (unpublished thesis submitted towards fulfilment of requirements for Diploma, I.C.A.R., New Delhi.)

(14)	Seshavadhanulu, M.	(1963)	"Some Contributions to the Theory of Sampling, (unpublished thesis submitted towards fulfilment of the requirements for Diploma, I.C.A.R., New Delhi.)
(15)	Seal, K.C.	(1951)	"On errors of estimation in various types of double sampling procedures, Sankhya, 11, 125-144.
(16)	-	(1953)	"On certain extended cases of double sampling", Sankhya, 12, 357-362.
(17)	Singh, D.	(1959)	"Paper read at the 12th Annual Meeting of the Indian Society of Agricultural Statistics held at Gwalior.
(18)	Singh, B.D.	(1962)	"Use of double sampling in repeated surveys", (unpublished thesis submitte-d towards fulfilment of requirements for Diploma, I.C.A.R., New Delhi.)
(19)	Singh, D., and Singh, B.D.	(1965)	"Some contributions to two-phase sampling". The Aus. J. of Stat. 1965 Vol. 7, No.2 45-47.
(20)	Sukhatme, P.V.	(1953)	"Sampling Theory of Surveys with applications", The Iowa State College Press, Ames, Iowa, U.S.A.
(21)	Sukhatme, B.V. and Koshal, R.S.	(1959)	"A contribution to double sampling", Jr. Ind. Soc. Agri. Stat., 11, 128-144.
(22)	Tikkiwal, B.D.	(1951)	"Theory of Successive Sampling", (unpublished thesis submitted towards partial fulfilment of requirements for Diploma, I.C.A.R., New Delhi).
(23)	-	(1954)	"Theory of Successive Multiphase Sampling", Abstract, Ann. Math. Stat., 25.
(24)	-	(1955)	"On the Efficiency of Estimates in Successive Multiphase Sampling", Abstract, Ann. Math. Stat., 26.
(25)	-	(1955)	"Multiphase Sampling on Successive Occasions", Ph.D. Thesis, Faculty of North Carolina State Colleges Raleigh, N.C., U.S.A.
(26)	-	(1956)	"A further contribution to the Theory of Unistage Sampling on Successive Occasions", J. Ind. Soc. Agri. Stat., 8.

(27) Tikkiwal, B.D. (1956) "An Application of the Theory of Multiphase Sampling on Successive Occasions to Surveys of Livestock Marketing", J. Karnatak Uni., 1.

(28) - (1958) "Theory of Successive Two Stage Sampling", Abstract, Ann. Math. Stat., 29.

(29) - (1964) "A note on Two Stage Sampling on Successive Occasions", Sankhya, Series A, 26.

(30) Yates, F. (1960) "Sampling methods for Censuses and Surveys" London: Charles Griffin & Co.