SOME ASPECTS OF SUCCESSIVE SAMPLING IN MULTISTAGE DESIGNS

By

O.P.Kathuria

Dissertation submitted in partial fulfillment of
the requirements for the award of Diploma
in Agricultural & Animal Husbandry
Statistics of the Institute of
Agricultural Research
Statistics
New Delhi.
1959.

ACKNOWLEDGEMENT

I have pleasure in expressing my deepest sense of gratitude to Shri D. Singh, Senior Research Statistician, 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 also due to the Institute of Agricultural Research Statistics for the facilities provided to me during the course of investigation.

O.P. Kathuria
16.4.60.

(O.P. Kathuria)

C O N T E N T S

Some Aspects Of Successive Sampling In Multistage Designs.

I N T R O D U C T I O N

Usually the results of a sample survey are useful for the occasion when the survey is conducted. A population which does not change with time presents no serious problem. 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 fertiliser use, or the number of unemployed persons in a country, any such survey is of limited use unless it has been repeated frequently at regular intervals of time. The duration of such a survey if it is undertaken, and the interval to be taken between two enquiries in such a survey will, of course, depend upon the type of the surveyed material, the information that is to be collected from it and the expenditure that has been sanctioned in conducting such a survey; this last aspect is a restrictive one as such the already dwindling resources of a country may not warrant a costlier survey to be undertaken. The importance of repeated surveys of the same population may be emphasized on other considerations also. Since dynamic changes are being brought about in the economy of our country, it would be instructive to take stock of the changes brought about at the end of each plan period, before we step into the next one.

Thus for a dynamic population, a sampler may be asked to present statistical estimates for the changes taking place in it from time to time, in respect of its various characters. He may be asked to give estimates for the average value of a character for the particular period under consideration. If, for instance, the population is subject to monthly variation, he may be required to give estimates for the population character for the current month and to compare it with the corresponding value in the preceeding month. Similarly if the population is subject to seasonal variations estimates may be required of the changes brought about by the seasonal effects.

Having made sure the objective of a survey the sampler would embark upon to adopt a suitable frame or design for the survey. Naturally his choice would be one among the various designs available. After the first round of enquiry is over, he may not like to choose the same sampling units for the second round as well, unless the units are extremely variable with time. If the sampling units are not so variable, a resurvey of the same units may fail to give any additional information of particular interest, rather much of time and resources may be lost in an endeavour to have such a survey. Moreover as Yates (1949) puts it a repeated resurvey of the same units may result in modification of these units relative to the rest of the population. He asserts the point by giving an example that in a survey of agricultural practice, visits to farms may result in the farmers concerned, improving their practice through the advice from the investigators; an advice which when asked for can scarcely be refused.

To select a fresh sample independently on each occasion, one will be confronted with a number of difficulties related to field operation. One is that if the enumerator is new to the place of enquiry, he will not get full co-operation of the local population in executing his work efficiently and secondly the whole procedure of preparing the basic frame, tabulating and listing of sampling units shall have to be repeated again which will mean consumption of more time and more travel to be undertaken and hence more cost per unit included in the sample.

Retaining some of the units from the previous enquiry and supplementing them with a sample selected a fresh from the population at each time seems to be an effective policy to be adopted for the field operation. There are some administrative advantages also in introducing the new units occasionally in fractions. Thus, for instance, in a survey involving tehsils and villages within tehsils as our sampling units, if we want to take a fresh sample of tehsils in the second enquiry, it may take at least a day more for every new tehsil included in the sample, for preparing the basic frame for selection of the sample from it and about three or four days more for enumerating the households within each selected village from this new tehsil. Therefore, it is desirable to have only some of the new units come into the sample at any time so as to spread the over-work over a number of enquiries.

In a survey involving a single stage random sample design partial replacement of sampling units presents no serious problem. But for a multi-stage sampling design the things are not so simple. Thus for a two-stage design, what fraction of primary sampling units should be selected afresh and what fraction to be retained on the second occasion, and if we have decided to retain all the psu's, whether all the second stage units within them should be replaced or they should be partially replaced and partially retained - these are the various problems with which a sampler is confronted.

While a good amount of work involving single-stage sampling design has been done by a number of workers, such as by Jessen, Yates, Patterson and Tikkiwal, practically no study has been made of the problem requiring more complex designs and to correlate the results with some actual survey. Secondly they have left off a more important aspect viz. the cost involved in a survey. Only Jessen's investigations are complete in this respect.

It would be opportune to outline some of the fields where the study made in the following pages can have a possible application. Schemes have been conducted by the Indian Council of Agricultural Research to assess and put forth reliable estimates for the milk-yield of cows and buffaloes, their breedwise distribution, feeding and other management practices in the Punjab, U.P., Bombay and Andhra Pradesh States. The survey conducted in Punjab has been dealt with in greater detail later on; the object here is to study in what form can the device of partial replacement of units be applied to the subsequent surveys. Since the successive sampling units are tehsils and villages, it may be possible to retain only some of tehsils for the second round of enquiry and complete the sample by taking some of the tehsils afresh. In that way a larger number of tehsils can be included in the survey. Since in the Punjab survey all the tehsils have been retained and villages within each tehsil taken afresh on each season it would be worthwhile to study whether we can replace a suitable fraction of tehsils and retain only some of the villages within each tehsil that has been retained and select some of them afresh so that the objective can be best achieved without any undue increase in the cost. If such a scheme is feasible, the objective is to find the number of tehsils and the number of villages within each tehsil that should be included in the sample on each occasion so that the estimate for the character under study can be built up with a precision that the given resources would permit.

A similar sampling plan may be adopted for the surveys which are repeated annually for estimating the yield of agricultural crops in the country. Mention may be made of a pilot sample survey for estimating the production and area under the cultivation of coconut and arecanut crops in Assam, Madras, Mysore, Andhra Pradesh, Bombay and Kerala States. Another scheme to assess the production of orange is being conducted in the Bombay State. The point of interest in the Coconut and Arecanut crop surveys is that for the second year of enquiry only a fraction of 20 to 25 percent of the villages sampled in the first year is to be retained this being supplemented by a fresh sample of villages. Thus as large as 75 to 80 percent replacement has been envisaged for the second and subsequent years of enquiry. As for the villages which have been retained it has been decided that the same set of palms should be taken for harvesting which were taken in the previous enquiry. The fresh sample of villages is to be taken from the entire population of villages including those also which have been retained; and if any of these retained villages happens to be selected again, it is not to be rejected but a double sample of gardens or trees is to be drawn from it for collecting information. Similar sampling plan for the second year has been adopted for the survey of oranges also. By adopting the scheme as given above, for the second year, improved estimates of production for these crops can be obtained.

To extend the field further mention may be made of the survey conducted in 1953 by the National Sample Survey to assess the employment and unemployment situation in the country. The best way to make a study of the improvements brought about by the introduction of five-year plans is to have an appraisal of the employment situation in the country at the end of each plan-period. Any design adopted for such a survey would essentially consist of towns and villages etc. to be taken as the sampling units. A suitable fraction of these may be replaced for subsequent surveys.

Perhaps the most realistic application of the replacement policy to a survey has been illustrated by Jessen in his "Statistical Investigations of a Sample Survey for Obtaining Farm Facts" in the Iowa State, U.S.A. In this study he shows that 'matched' samples were 25 to over 20 times as efficient as independent samples on each of the two occasions depending upon the type of the item for which information was required. His investigations also included a cost function appropriate for the survey and he has obtained optimum size of the sampling units for the given cost situations.

In passing mention may also be made of a sample survey being conducted in the U.S.A for the last ten years to estimate the annual dollar volume of sales of retail stores in that country. Estimates are also given for the percentage change in the volume of sales from month to month in the same year as also for the same month a year later. Replacement policy may effectively be adopted in such a survey.

REVIEW OF LITERATURE

Sampling on successive occasions with partial replacement of units was first studied by Jessen (1942) for unistage units only. Jessen's work was confined to two occasions only wherein the information obtained from the first occasion was utilised to build an estimate for the average on the second occasion. He considered two independent estimates for the current occasion; one sample consisted of units which were common to both the occasions, the information collected from these units on the preceeding occasion served as supplementary information for the estimate on the current occasion. Another estimate was the sample mean based on the units selected afresh on the current occasion. Both served as independent estimates of the population mean on the current occasion. The two were combined together with weights which minimised the variance of this new estimate and these weights were proportional to the reciprocals of their variances. An optimum value for the proportion of units which were common to the two occasions was also determined.

Yates (1949), however, has taken a liberal view of the situation. He contends that for estimating the value of the population mean on two successive occasions, it is more suitable to treat each occasion separately, following whatever method of estimation is appropriate to the sample obtained on that occasion, regardless of the values obtained on the other occasion. For two occasions, he has considered a subsample of the original sample as also a sample with some of the units retained from the previous occasion and some taken afresh on the second occasion. When a subsample of the original sample is taken, an estimate of change will be obtained from units included in the subsample only. An estimate of the population mean on the second occasion may be obtained either by adding this estimated change to the overall estimate on the first occasion, or a regression estimate may be built with these units by using the sample values on the first occasion as supplementary information. Yates has also extended his results to h occasions ($h > 2$) and has built an estimate for the population mean on the hth occasion by taking into account the results upto and including the (h - 1)th occasion. Subject to certain limitations, the estimates Y_h and Y_{h-1} for h and (h-1)th occasions are related by

$$Y_h = (1-\phi_h)\left\{\bar{y}'_h + \rho\,(\bar{Y}_{h-1} - \bar{y}'_{h-1})\right\} + \phi_h\,\bar{y}''_h$$

where single dashes denote units common to occasions h and (h-1), the mean on the earlier occasion is indicated by the squre brackets and double dashes units occuring on occasion h only.

The limitations that have been imposed are that a given fraction of units is replaced on each occasion; the variability on different occasions is constant i.e. $\sigma_h^2 = \sigma_{h-1}^2 = \sigma^{*2}$ for all h; the correlation ρ between the units on the successive occasions is constant; and that correlation between the units occasions two apart is ρ^2 that between the units occasions three apart is ρ^3 etc. The value of ϕ_h varies from occasion to occasion and it depends upon the values of ρ and the fraction

μ replaced on each occasion. It acts as a weight combining a regression estimate and an another unbiased estimate. Expression has also been obtained for the change $\bar{\bar{Y}}_h - \bar{\bar{Y}}_{h-1}$ between the estimates on the hth and (h-1)th occasions. For different values of ρ and for $\mu = 1/2$ and $\mu = 1/3$ tables have been prepared for the efficiency of the estimates of mean and change relative to the overall mean and the difference of the overall means for (a) two and (b) h occasions. Relative efficiency of a subsample has also been discussed.

Patterson's (1950) approach to the problem is slightly more general. He first builds an estimate as a suitable linear function of a set of variates and then develops a set of conditions for this estimate to be the most efficient. He then makes use of these conditions to get an efficient estimate of the mean on the hth occasion, which comes out to be the same as given by Yates. With this set of conditions he also establishes a recurrence relation between the weights ϕ_h and ϕ_{h-1} and a limiting value to $\phi_h = \phi$ when h is very large is given by

$$\phi \doteq \frac{-(1-\rho^2) + \sqrt{(1-\rho^2)\left[1-\rho^2(1-4\lambda\mu)\right]}}{2\lambda\rho^2}$$

where $\lambda + \mu = 1$.

When this limiting weight is used there is a slight increase in the variance of Y_h. A solution suggested by him is that the exact weights should be used for the first two occasions and thereafter, the limiting weights may be used in place of ϕ_3, ϕ_4 -------- etc. When the sample on the hth occasion has been taken, the information provided by it may be utilised to improve the estimate for the $(h-k)$th occasion for $k > 1$. If $_hY_{h-1}$ be this refined estimate for the $(h-1)$th occasion then its expression is given by

$$_hY_{h-1} = Y_{h-1} - \rho\,\phi_{h-1}(Y_h - \bar{y}_h'')$$

Patterson shows that the change $Y_h - {}_hY_{h-1}$ is more efficient than the one given by Yates. However, for $h = 2$ this new estimate of change has been considered by Yates also.

Tikkiwal's (1953) work follows the same lines, as set by Yates and Patterson; only difference is that the pattern of correlation set by him is slightly more general than the one followed by his predecessors. While the correlation between units taken on different occasions has been allowed to vary, that between units two or more than two occasions apart has been taken to be equal to the product of correlations between units on all pairs of consecutive occasions formed by these. In case the sample size and the correlations are constant on all the occasions, he shows that with the limiting ϕ, the replacement to be affected on different occasions is 50 percent. This limiting value is attained from above, meaning thereby that under the conditions imposed the replacement fraction is always $\geq 1/2$.

When the correlation and regression coefficients are not known in advance but are estimated from the sample values the weights ϕ_h will again be changed and the variability of the correlation and regression coefficients as computed from the samples shall have to be taken into account. This is particularly important when the number of units common to the both occasions is very small and the correlation coefficient is calculated from these units. The case has been discussed for two occasions by Jessen and for h occasions by Narain (1953).

In another paper (1956) Tikkiwal has shown that when the correlation and regression coefficients are estimated from the common units between two consecutive occasions, Y_h is still a consistent estimator of the population mean μ_h on the hth occasion and its bias tends to zero with increasing sample sizes on h occasions. Its variance will in general be greater than the variance of the estimator where the correlations are known in advance and the weights (ϕ_h) themselves become functions of parameters to be estimated from the sample.

A study of partial replacement with multistage units has been made by D. Singh (1959).Only two stage units on two occasions have been taken. A fraction of psu's has been taken to be common on both the occasions. Expression has been obtained for the estimate of the mean and its variance. When a survey has been repeated on different intervals of time over a given period, an estimate has been obtained for the mean for the entire period or for any particular interval. It has been observed that for a survey repeated at three equal intervals over a given period, the procedure with independent samples at each interval will be more efficient than the one when some of the units sampled on the first interval are repeated again on the subsequent intervals. Partial replacement of units can only be effective if the correlation coefficients between the values of units on the different occasions is negative.

SCHEME FOR THE STUDY OF MILK YIELDS, BREEDS AND FEEDING AND MANAGEMENT PRACTICES OF CATTLE AND BUFFALOES IN PUNJAB (1956-57)

****

A pilot scheme to obtain reliable estimates of the milk yield and collect information on feeds and other management practices of cattle and buffaloes was conducted in the Punjab State in 1956-57. The entire area excluding some hilly tracts was brought under the survey. The field work of the survey was spread over all the three seasons of the year, March to June, July to September and October to February constituting the summer, rainy and winter seasons respectively.

The sampling plan adopted for the survey was one of stratified multistage random sampling. The entire area was divided into three composite geographical zones, the Northern, Central and the Southern and these constituted the three strata. Tehsils and villages within tehsils constituted the psu's and ssu's respectively. The sample consisted of 15 tehsils, 5 from each stratum. In the summer season a random sample of 4 villages was selected from each of the first two selected tehsils in a zone. From each of the remaining three tehsils, a sample of two villages and one town with two wards from each town were selected, the quantum of work in a ward being equal to that in a village. An enumerator was appointed for each tehsil who made one complete round of all the selected villages and towns in a month, spending about a week in each village or town. Since the summer season consists of four months, the field work consisted of 4 rounds of each selected village or ward. The plan of work adopted in the rainy and winter seasons was slightly different from that adopted in the summer season. Two random clusters of three villages each or one cluster and two towns with one ward from each town, was the plan adopted for these two seasons corresponding to 4 villages or 2 villages and one town with two wards from it - as in the summer season. The time to be spent in villages in each cluster or in the town was determined beforehand.

The salient features of this survey were the collection of detailed data on milk yield and different feeds given to the animals, their breed-wise distribution, age, sex and order and stage of lactation of animals in milk. Information was also collected on the veterinary facilities available in the towns and villages to set them on proper footing.

The staff for the field work consisted of, besides the fifteen enumerators, three supervisory-cum-relieving enumerators, three inspectors and one field officer. The statistical staff consisted of one Assistant Statistician, one Statistical Asstt., and four senior computors.

The total cost of the scheme was of the order of Rs.1.03 lakhs, of which the expenditure estimated on the field staff was Rs. 63,000/- and that on the statistical staff was Rs.40,000/-.

Estimates for the components of variance for the various stages of sampling units have been obtained and these are used in the present investigation. The objective is to determine the various components of cost involved in the field work of the survey and to build a suitable cost function for the scheme. Only two stages of units viz. the tehsils and villages within tehsils have been taken and therefore the cost function involves components of cost for these two stages only. Although in the actual survey same tehsils have been taken in all the three seasons and only the villages within each tehsil are taken afresh in each season, the object of the present investigation is to obtain the optimum number of psu's and ssu's within each psu which should constitute the sample for a given cost function. Thus assuming that a fixed proportion of tehsils is replaced in each season, optimum values for the number of tehsils and villages within each tehsil have been obtained which would best estimate the average milk yield of animals in the second season of enquiry subject to the cost function built up for the survey. Any change in the values of the components of cost which are involved in the cost function or any change in the estimates for the variance components (which will necessarily be different for different surveys) would alter the optimum value for the size of the sample. Tables have thus been prepared for obtaining the optimum values for the number of psu's and the ssu's to be taken in the sample for different values of the variance components and under different cost patterns.

Sampling on two occasions, a fraction p of the first-stage units being retained:- Let us first consider the case when sampling is done on two occasions only with a two-stage sampling design. For simplicity we shall assume sampling with replacement both for the psu's and ssu's. Let N and M be the number of psu's and ssu's in the population and suppose that a simple random sample of n psu's is drawn for the first occasion and that out of the ith psu selected, we select m_i ssu's, so that the sample on the first occasion consists of $\sum_i^n m_i = m_0$ sampling units. Suppose further that we retain a simple random sample of np psu's for the sample on the second occasion, and supplement it by qn independently selected psu's, where $p + q = 1$, so that $\sum_i^{np} m_i = m_{01}$ units are common to the two occasions and $\sum_i^{nq} m_i = m_{02}$ units are selected afresh on the second occasion, the total sample size for the second occasion being maintained the same as that for the first occasion. Practical considerations in an actual field survey necessitate the size of the sampling units to be the same on all occasions. For, any change in the sample size on a subsequent occasion may require a change in the field staff also, and the field work may also be disturbed.

Now let:

$\bar{x}_1'$ = mean on the first occasion, based on the $\sum_i^{np} m_i$ units which are also common to the second occasion.

$\bar{y}_2'$ = mean on the second occasion associated with such units.

$\bar{x}_1''$ = mean on the first occasion based on $\sum_i^{nq} m_i$ units which are not common with the second occasion.

and $\bar{y}_2''$ = mean on the second occasion based on $\sum_i^{nq} m_i$ units which are selected afresh.

The above estimates are defined as

$$\bar{y}_2' = 1/m_{o1} \sum_{i=1}^{np} \sum_{j=1}^{m_i} y_{ij}$$

$$= 1/m_{o1} \sum_{i}^{np} m_i \, \bar{y}_{i(m_i)}$$

$$\bar{y}_2'' = 1/m_{o2} \sum_{i}^{nq} \sum_{j=1}^{m_i} y_{ij}$$

$$= 1/m_{o2} \sum_{i=1}^{nq} m_i \, \bar{y}_{i(m_i)}$$

Similarly for the estimates $\bar{x}_1'$, $\bar{x}_1''$ on the first occasion also.

$\bar{y}_2'$, $\bar{y}_2''$ provide unbiased estimates for the mean $\bar{Y}_2$ on the second occasion and so also $\bar{x}_1'$, $\bar{x}_1''$ for the population mean $\bar{Y}_1$ on the first occasion.

Estimate of mean:- We wish to estimate $\bar{Y}_2$ by a linear function of the form

$$\bar{y}_2 = a\,\bar{x}_1' + b\,\bar{x}_1'' + c\bar{y}_2' + d\,\bar{y}_2'' .$$

Obviously it is given by

$$\bar{y}_2 = a\,(\bar{x}_1'' - \bar{x}_1') + c\bar{y}_2' + (1-c)\,\bar{y}_2'' \quad \ldots\ldots \quad (1)$$

where a and c are obtained by minimising $V(\bar{y}_2)$ with respect to these quantities.

Now $$V(\bar{y}_2) = a^2 \left[V(\bar{x}_1') + V(\bar{x}_1'') \right] + c^2 V(\bar{y}_2') + (1-c)^2 V(\bar{y}_2'') + 2\,ac\,\mathrm{Cov}(\bar{y}_2', \bar{x}_1')$$

the remaining product terms do not come into picture under the type of replacement adopted.

Following Sukhatme (1953) we write

$$V(\bar{x}_1') = 1/m_{o1}\; S_w^2 + \frac{\sum_{i=1}^{np} m_i^2}{m_{o1}^2} S_b^2$$

$$= V(\bar{y}_2) = A \text{ (say)} \quad \ldots\ldots \quad (2)$$

$$V(\bar{x}_1'') = \frac{S_w^2}{m_{o2}} + \frac{\sum_{i=1}^{nq} m_i^2}{m_{o2}^2} S_b^2$$

$$= V(\bar{y}_2'') = B \text{ (say)} \quad \ldots\ldots\ldots \quad (3)$$

Assumptions made for this are that the variability within each psu is constant whether on the first or the second occasion.

i.e. $S_{ix}^2 \simeq S_w^2$ and $S_{iy}^2 \simeq S_w^2$ for all i and for all x and y and

$$S_{bx}^2 = S_{by}^2 \simeq S_b^2$$

Similarly we shall have

$$\text{Cov}(\bar{x}_1', \bar{y}_2') = \frac{\rho_2 S_w^2}{m_{o1}} + \frac{\rho_1 \sum_{i=1}^{np} m_i^2}{m_{o1}^2} S_b^2 = C \text{(say)}. \, (4)$$

where again it is assumed that

$$\rho_{ixy} = \rho_{wxy} \doteq \rho_2$$

$$\text{AND} \quad \rho_{bxy} \doteq \rho_1$$

ρ_1 and ρ_2 are defined as

$$\rho_1 = \rho_{bxy} = \frac{\sum_{i=1}^{N} (\bar{y}_{i.} - \bar{\bar{y}})(\bar{x}_{i.} - \bar{\bar{X}})}{\sqrt{\sum_{i=1}^{N} (\bar{y}_{i.} - \bar{\bar{Y}})^2 \sum_{i=1}^{N} (\bar{x}_{i.} - \bar{\bar{X}})^2}} \quad \ldots\ldots (5)$$

$$\text{and } \rho_2 = \rho_{wxy} = \frac{\sum_{i}^{N} \sum_{j=1}^{M_i} (\bar{y}_{ij(m_i)} - \bar{y}_{i.})(\bar{x}_{ij(m_i)} - \bar{x}_{i.})}{\sqrt{\sum_{i=1}^{N} \sum_{j=1}^{M_i} (\bar{y}_{ij(m_i)} - \bar{y}_{i.})^2 \sum_{i=1}^{N} \sum_{j=1}^{M_i} (\bar{x}_{ij(m_i)} - \bar{x}_{i.})^2}} \quad \ldots\ldots (6)$$

Again it is assumed that sampling is done with replacement, both among psu's and ssu's. We wish to choose the values for a and c which minimise $V(\bar{y}_2)$.

This gives

$$c = \frac{\left\{V(\bar{x}_1') + V(\bar{x}_1'')\right\} \quad V(\bar{y}_2'')}{\left\{V(\bar{y}_2') + V(\bar{y}_2'')\right\}\left\{V(\bar{x}_1') + V(\bar{x}_1'')\right\} - \left\{Cov(\bar{y}_2', \bar{x}_1')\right\}^2}$$

and

$$a = \frac{V(\bar{y}_2'') \quad Cov(\bar{y}_2', \bar{x}_1')}{\left\{V(\bar{y}_2') + V(\bar{y}_2'')\right\}\left\{V(\bar{x}_1') + V(\bar{x}_1'')\right\} - \left\{Cov(\bar{y}_2', \bar{x}_1')\right\}^2}$$

Substituting the values of the variances in terms of A, B and C we have

$$c = \frac{(A + B)\, B}{\left[(A + B)^2 - C^2\right]} \qquad \ldots\ldots (7)$$

and

$$a = \frac{BC}{\left[(A + B)^2 - C^2\right]} \qquad \ldots\ldots (8)$$

This, therefore, gives the required estimate as

$$\bar{y}_2 = \frac{(A + B)\, B}{\left[(A+B)^2 - C^2\right]} \left\{ \bar{y}_2' + \frac{C}{q\,(A+B)} (\bar{X}_1 - \bar{x}_1') \right\} + \frac{\left[A\,(A + B) - C^2\right]}{\left[(A + B)^2 - C^2\right]} \bar{y}_2'' \quad \ldots (9)$$

which may be seen to be the weighted average of the two estimates

$$\left\{ \bar{y}_2' + \frac{C}{q\,(A + B)} (\bar{X}_1 - \bar{x}_1') \right\} \text{ and } \bar{y}_2''$$

the weights being given by

$$\frac{(A+B)\,B}{\left[(A+B)^2 - C^2\right]} \doteq 1 - \psi_2$$

and

$$\frac{\left[A\,(A+B) - C^2\right]}{\left[(A+B)^2 - C^2\right]} = \psi_2 \quad \text{(say)} \quad \ldots\ldots(10)$$

Variance of the estimate is given by

$$V(\bar{y}_2) = \frac{B\left[A\,(A+B) - C^2\right]}{\left[(A+B)^2 - C^2\right]}$$

Since A, B and C are functions of m_i's, it would be of interest to find the size of each psu, for which $V(\bar{y}_2)$ will be minimum. But it will not give any idea about the optimum number of psu's which should be replaced on the second occasion. For this, we shall assume that $m_i = m$ for all i. This then gives

$$A = \frac{1}{np}\left(S_b^2 + \frac{S_w^2}{m}\right) = \frac{\alpha}{np} \quad \text{(say)}$$

Similarly $B = \dfrac{\alpha}{nq}$

and $$C = \frac{1}{np}\left(\rho_1 S_b^2 + \rho_2 \frac{S_w^2}{m}\right) = \frac{\gamma}{np} \quad \text{(say)}$$

Therefore

$$\bar{y}_2 = \frac{p}{\left(1 - \frac{\gamma^2}{\alpha^2} q^2\right)}\left\{ y_2' + \frac{\gamma}{\alpha}\left(\bar{X}_1 - \bar{x}_1'\right)\right\} + \frac{q\left(1 - \frac{\gamma^2}{\alpha^2} q\right)}{\left(1 - \frac{\gamma^2}{\alpha^2} q^2\right)}\,\bar{y}_2'' \quad \ldots(11)$$

and

$$V(\bar{y}_2) = \frac{\alpha\left[\alpha^2 - \gamma^2 q\right]}{n\left[\alpha^2 - \gamma^2 q^2\right]} \quad \ldots\ldots\ldots(12)$$

and

$$\psi_2 = q \cdot \frac{\left[\alpha^2 - \gamma^2 q\right]}{\left[\alpha^2 - \gamma^2 q^2\right]} \quad \ldots\ldots\ldots(13)$$

It is now possible to get the optimum value of the replacement fraction q, for which $V(\bar{y}_2)$ is minimum. This is given by

$$\frac{d}{dq} V(\bar{y}_2) = 0$$

Therefore $$q_{opt} = \frac{\alpha^2 - \alpha\sqrt{\alpha^2 - \gamma^2}}{\gamma^2} \quad \ldots\ldots(14)$$

and $$V_{opt}(\bar{y}_2) = \frac{1}{2n}\left(\alpha + \sqrt{\alpha^2 - \gamma^2}\right) \quad \ldots\ldots(15)$$

(12) shows that for q = 0 and for q = 1 i.e. for no replacement or for complete replacement of the units, the variance of the estimate has the same value $\frac{\alpha}{n}$.

Further, since $\gamma \leqslant \alpha$, it follows from (14) that the replacement fraction q should at least be as large as 1/2, i.e. not more than 50% of the units should be retained from the first occasion to the second and their percentage decreases steadily as γ increases.

We assume that ρ_1 and ρ_2 are both positive, for, if a steady improvement is taking place in the character under consideration, this will be reflected by the positive correlations between the psu's taken on the two occasions as well as between the ssu's within them.

We write $$\phi = \frac{S_w^2}{S_b^2}$$

Therefore $$\gamma/\alpha = \frac{\rho_1 + \rho_2\,\phi/m}{1 + \phi/m} \quad (16)$$

Table 1. gives for a series of values of ρ_1, ρ_2, ϕ and m, the optimum percentage of psu's which should be replaced and the relative gain in precision as compared to complete replacement. Values of ρ_1 have been taken from .50 to 1.0. For

$\rho_1 < .50$ the gains are very modest. ρ_2 varies from .10 to .90. ϕ ranges from .10 to 10.0 as different surveys may yield different values for this ratio; m has been given three values 3, 5 and 7 in order to see how far an increase in the size of the subsample brings about a change in the replacement of the psu's. Upper values in each bracket in the table denote the percent replacement of psu's between two occasions, while the lower values denote the percent gain in precision as compared to complete replacement.

From the table it can be seen that:-

i) as the correlations ρ_1 and ρ_2 between the psu's and the ssu's increase, the replacement percentage also increases; i.e. a larger proportion of new units should be added to the sample on the second occasion. Consequently the gain in precision of the estimate on the second occasion also increases. Of course it may be prohibitive from cost considerations to add larger number of new units to the sample.

ii) the replacement percentage and the gain in precision of the estimate, increase more rapidly when ρ_1 increases than when ρ_2 increases. This means that the correlation between the psu's is more important in bringing about a larger gain in precision than that between the corresponding second stage units within them. This is evident otherwise also - when the psu's in the sample tend to be alike in respect of any particular character under consideration, only a few of them need be retained for the subsequent occasion and more of new psu's should be brought into the sample.

iii) for a fixed ρ_1 so long as ρ_2 is less than ρ_1 ;

Table 1 Values for the Optimum replacement percentage and the percentage of gain in precision of the estimate relative to complete replacement for a series of values of ρ_1, ρ_2, ϕ and m, Opt. q and Opt. $V(\bar{y}_2)$ being given by (14) and (15).

ρ_1	$\rho_2 \rightarrow$ / $\downarrow\phi$	.10	.40	.50	.60	.80	.90	.10	.40	.50	.60	.80	.90	.10	.40	.50	.60	.80	.90
		m = 3						5						7					
.50	.10	$\binom{53}{7}$	$\binom{53}{7}$	$\binom{53}{7}$	$\binom{53}{7}$	$\binom{54}{7}$	$\binom{54}{8}$	$\binom{53}{7}$	$\binom{53}{7}$	$\binom{53}{7}$	$\binom{53}{7}$	$\binom{54}{7}$	$\binom{54}{7}$	$\binom{53}{7}$	$\binom{53}{7}$	$\binom{53}{7}$	$\binom{53}{7}$	$\binom{54}{7}$	$\binom{54}{7}$
.50	.50	$\binom{53}{7}$	$\binom{53}{7}$	$\binom{53}{7}$	$\binom{54}{7}$	$\binom{54}{8}$	$\binom{56}{9}$	$\binom{53}{6}$	$\binom{53}{7}$	$\binom{53}{7}$	$\binom{53}{7}$	$\binom{54}{8}$	$\binom{54}{8}$	$\binom{53}{6}$	$\binom{53}{7}$	$\binom{53}{7}$	$\binom{53}{7}$	$\binom{54}{8}$	$\binom{54}{8}$
.50	1.0	$\binom{52}{4}$	$\binom{53}{6}$	$\binom{53}{7}$	$\binom{54}{8}$	$\binom{55}{10}$	$\binom{55}{11}$	$\binom{52}{5}$	$\binom{53}{6}$	$\binom{53}{7}$	$\binom{54}{8}$	$\binom{54}{9}$	$\binom{55}{10}$	$\binom{53}{5}$	$\binom{53}{7}$	$\binom{53}{7}$	$\binom{54}{7}$	$\binom{54}{9}$	$\binom{54}{9}$
.50	4.0	$\binom{51}{2}$	$\binom{53}{5}$	$\binom{53}{7}$	$\binom{55}{9}$	$\binom{57}{15}$	$\binom{59}{17}$	$\binom{51}{2}$	$\binom{53}{6}$	$\binom{53}{7}$	$\binom{54}{9}$	$\binom{57}{14}$	$\binom{58}{16}$	$\binom{52}{4}$	$\binom{53}{6}$	$\binom{53}{7}$	$\binom{54}{8}$	$\binom{56}{12}$	$\binom{57}{14}$
.50	10.0	$\binom{50}{1}$	$\binom{52}{4}$	$\binom{53}{7}$	$\binom{56}{12}$	$\binom{59}{18}$	$\binom{63}{26}$	$\binom{50}{1}$	$\binom{52}{5}$	$\binom{53}{7}$	$\binom{55}{10}$	$\binom{58}{17}$	$\binom{61}{22}$	$\binom{51}{2}$	$\binom{53}{5}$	$\binom{53}{7}$	$\binom{55}{9}$	$\binom{58}{16}$	$\binom{60}{21}$
.70	.10	$\binom{58}{16}$	$\binom{58}{16}$	$\binom{58}{16}$	$\binom{58}{18}$	$\binom{58}{17}$	$\binom{58}{17}$	$\binom{58}{16}$	$\binom{58}{16}$	$\binom{58}{17}$	$\binom{58}{17}$	$\binom{58}{17}$	$\binom{58}{17}$	$\binom{58}{16}$	$\binom{58}{17}$	$\binom{58}{17}$	$\binom{58}{17}$	$\binom{58}{17}$	$\binom{58}{17}$
.70	.50	$\binom{56}{12}$	$\binom{57}{15}$	$\binom{58}{16}$	$\binom{58}{16}$	$\binom{59}{18}$	$\binom{59}{18}$	$\binom{57}{14}$	$\binom{57}{15}$	$\binom{58}{16}$	$\binom{58}{16}$	$\binom{59}{18}$	$\binom{59}{18}$	$\binom{57}{15}$	$\binom{58}{16}$	$\binom{58}{17}$	$\binom{58}{17}$	$\binom{58}{17}$	$\binom{59}{18}$
.70	1.0	$\binom{54}{9}$	$\binom{56}{12}$	$\binom{57}{14}$	$\binom{58}{16}$	$\binom{60}{20}$	$\binom{60}{21}$	$\binom{56}{11}$	$\binom{56}{12}$	$\binom{57}{14}$	$\binom{58}{16}$	$\binom{59}{18}$	$\binom{59}{18}$	$\binom{57}{15}$	$\binom{58}{16}$	$\binom{58}{16}$	$\binom{58}{17}$	$\binom{59}{18}$	$\binom{59}{19}$
.70	4.0	$\binom{52}{4}$	$\binom{55}{10}$	$\binom{56}{13}$	$\binom{57}{15}$	$\binom{61}{21}$	$\binom{62}{23}$	$\binom{52}{5}$	$\binom{55}{10}$	$\binom{56}{13}$	$\binom{57}{15}$	$\binom{60}{20}$	$\binom{62}{23}$	$\binom{53}{6}$	$\binom{56}{12}$	$\binom{57}{14}$	$\binom{57}{15}$	$\binom{60}{20}$	$\binom{61}{22}$
.70	10.0	$\binom{51}{2}$	$\binom{53}{7}$	$\binom{55}{10}$	$\binom{56}{12}$	$\binom{62}{23}$	$\binom{68}{32}$	$\binom{51}{2}$	$\binom{53}{7}$	$\binom{55}{10}$	$\binom{56}{12}$	$\binom{61}{22}$	$\binom{64}{28}$	$\binom{52}{4}$	$\binom{54}{8}$	$\binom{55}{11}$	$\binom{56}{13}$	$\binom{61}{22}$	$\binom{64}{27}$

Contd.....on next page.....

ρ_1	ρ_2 / Φ	.10	.40	.50	.60	.80	.90	.10	.40	.50	.60	.80	.90	.10	.40	.50	.60	.80	.90
	.10	$\binom{68}{55}$	$\binom{68}{57}$	$\binom{68}{57}$	$\binom{68}{57}$	$\binom{69}{59}$	$\binom{69}{59}$	$\binom{68}{55}$	$\binom{68}{57}$	$\binom{68}{57}$	$\binom{68}{57}$	$\binom{69}{59}$	$\binom{69}{59}$	$\binom{68}{57}$	$\binom{68}{57}$	$\binom{69}{58}$	$\binom{69}{59}$	$\binom{69}{59}$	$\binom{69}{59}$
.90	.50	$\binom{64}{28}$	$\binom{65}{51}$	$\binom{66}{52}$	$\binom{67}{54}$	$\binom{69}{57}$	$\binom{69}{59}$	$\binom{64}{28}$	$\binom{65}{51}$	$\binom{66}{52}$	$\binom{67}{54}$	$\binom{69}{57}$	$\binom{69}{59}$	$\binom{65}{51}$	$\binom{67}{54}$	$\binom{67}{54}$	$\binom{68}{55}$	$\binom{69}{59}$	$\binom{69}{59}$
	1.0	$\binom{58}{17}$	$\binom{61}{23}$	$\binom{62}{24}$	$\binom{65}{27}$	$\binom{68}{55}$	$\binom{69}{59}$	$\binom{61}{22}$	$\binom{64}{27}$	$\binom{64}{28}$	$\binom{65}{51}$	$\binom{68}{55}$	$\binom{69}{59}$	$\binom{65}{25}$	$\binom{65}{50}$	$\binom{65}{51}$	$\binom{66}{52}$	$\binom{68}{57}$	$\binom{69}{59}$
	4.0	$\binom{55}{5}$	$\binom{56}{12}$	$\binom{57}{15}$	$\binom{59}{18}$	$\binom{66}{32}$	$\binom{69}{59}$	$\binom{54}{9}$	$\binom{58}{16}$	$\binom{59}{19}$	$\binom{61}{22}$	$\binom{68}{52}$	$\binom{69}{59}$	56 $\binom{55}{12}$	5 $\binom{59}{18}$	$\binom{60}{21}$	$\binom{66}{24}$ 2	$\binom{68}{36}$	$\binom{69}{59}$
	10.0	$\binom{51}{2}$	$\binom{54}{8}$	$\binom{56}{11}$	$\binom{57}{15}$	$\binom{64}{27}$	$\binom{69}{59}$	$\binom{52}{4}$	$\binom{55}{10}$	$\binom{56}{15}$	$\binom{58}{17}$	$\binom{64}{28}$	$\binom{69}{59}$	$\binom{52}{5}$	$\binom{56}{12}$	$\binom{57}{14}$	$\binom{58}{17}$	$\binom{65}{50}$	$\binom{69}{59}$
	.10	$\binom{81}{61}$	$\binom{85}{76}$ 67	$\binom{85}{70}$	$\binom{87}{74}$	$\binom{88}{76}$	$\binom{94}{88}$	$\binom{85}{67}$	$\binom{86}{72}$	$\binom{87}{74}$	$\binom{88}{75}$	$\binom{91}{82}$	$\binom{94}{88}$	$\binom{87}{74}$	$\binom{87}{74}$	$\binom{88}{75}$	$\binom{88}{76}$	$\binom{93}{86}$	$\binom{96}{92}$
	.50	$\binom{67}{54}$	$\binom{72}{42}$	$\binom{73}{45}$	$\binom{74}{48}$	$\binom{81}{61}$	$\binom{87}{74}$	$\binom{72}{44}$	$\binom{74}{48}$	$\binom{76}{52}$	$\binom{78}{56}$	$\binom{85}{67}$	$\binom{88}{75}$	$\binom{74}{48}$	$\binom{78}{56}$	$\binom{80}{59}$	$\binom{81}{61}$	$\binom{87}{74}$	$\binom{91}{82}$
1.0	1.0	$\binom{61}{25}$	$\binom{65}{51}$	$\binom{67}{55}$	$\binom{69}{59}$	$\binom{76}{51}$	$\binom{81}{62}$	$\binom{65}{21}$	$\binom{69}{59}$	$\binom{71}{42}$	$\binom{72}{45}$	$\binom{81}{61}$	$\binom{85}{67}$	$\binom{88}{57}$	$\binom{73}{45}$	$\binom{74}{48}$	$\binom{75}{50}$	$\binom{85}{67}$	$\binom{87}{74}$
	4.0	$\binom{55}{7}$	$\binom{57}{15}$	$\binom{59}{18}$	$\binom{61}{21}$	$\binom{68}{37}$	$\binom{75}{50}$	$\binom{56}{11}$	$\binom{59}{18}$	$\binom{62}{25}$	$\binom{64}{27}$	$\binom{71}{42}$	$\binom{78}{56}$	$\binom{57}{15}$	$\binom{82}{23}$	$\binom{64}{28}$	$\binom{66}{52}$	$\binom{73}{45}$	$\binom{78}{58}$
	10.0	$\binom{51}{2}$	$\binom{54}{9}$	$\binom{56}{15}$	$\binom{58}{17}$	$\binom{66}{51}$	$\binom{72}{44}$	$\binom{52}{6}$	$\binom{56}{11}$	$\binom{58}{15}$	$\binom{59}{18}$	$\binom{67}{24}$	$\binom{78}{46}$	$\binom{53}{6}$	$\binom{57}{14}$	5 $\binom{69}{18}$	$\binom{61}{21}$	$\binom{68}{55}$	$\binom{74}{48}$

the replacement percentage, and the gain in precision decrease as ϕ increases and they start increasing as soon as ρ_2 assumes value greater than that of ρ_1. This means that when the correlation between the ssu's is smaller than that between the corresponding psu's in the sample, then it is only when the variation between the ssu's is smaller than that between the psu's, shall we expect the precision of the estimate to increase relative to complete replacement; and if the correlation between the ssu's is larger than that between the psu's then we should expect the variation between the psu's to be larger than that between the corresponding ssu's within them.

iv) when ρ_2 is large as compared to ρ_1 the gain in precision is more when m = 3 than when m = 5 or 7. Thus for larger correlation between the ssu's in the sample, only a few of them need be taken in the sample on the subsequent o casion.

Thus when the psu's in a sample tend to be alike each other with respect to the character under study,only a few of them should be retained on the second occasion and a larger subsample may be taken from each selected psu. Similarly when the ssu's within each psu are alike each other, then a smaller subsample from each psu should be taken and a larger number of psu's may be included in the sample. The optimum size of the sample has been dealt with later on when the cost of the survey is also taken into consideration.

<u>Estimate of change:</u> The estimate of change may be of particular interest in order to apprise oneself of the effectiveness of any development scheme.

A simple estimate of change is

$$D = \bar{y}_2 - \bar{x}_1 = p\ (\bar{y}_2' - \bar{x}_1'\ ) + q\ (\bar{y}_2'' - \bar{x}_1''\ ) \quad \ldots.(17)$$

and $$\sigma_D^2 = \frac{2\ (\alpha - \gamma\ p\ )}{n}$$

A second estimate of change may be obtained by utilising the information provided by the units selected on both occasions. This new estimate would be similar to (17) but with different weights.

This may easily be obtained as

$$D_w = \frac{p\ \alpha}{(\alpha - \gamma\ q\ )} (\bar{y}_2' - \bar{x}_1'\ ) + \frac{q(\alpha - \gamma\ )}{(\alpha - \gamma\ q)} (\bar{y}_2'' - \bar{x}_1''\ ) \quad \ldots.(19)$$

and its variance is given by

$$\sigma_{D_w}^2 = \frac{2\ \alpha\ (\alpha - \gamma\ )}{n\ \ (\alpha - \gamma\ q)} \quad \ldots.(20)$$

Therefore

$$\frac{\sigma_D^2}{\sigma_{D_w}^2} = \frac{(1 - \gamma/\alpha + \gamma^2/\alpha^2\ pq)}{(\ 1 - \gamma/\alpha\ )} \quad \ldots.(21)$$

This indicates that $\sigma_{D_w}^2$ may be made considerably small as compared to σ_D^2 depending upon the value which the quantity $\dfrac{pq.\ \gamma^2/\alpha^2}{(1 - \gamma/\alpha\ )}$ takes, and this quantity will be large when ρ_1 and ρ_2 are large, in which case the second estimate may be made more efficient than the previous one.

Table 2. gives for a series of values of ρ_1, ρ_2, ϕ, m and q, the percentage efficiency of the estimate D_W as compared to D. This efficiency is given by the quantity

$$\left\{ \frac{\sigma_D^2}{\sigma_{D_W}^2} - 1 \right\} \times 100 \quad \ldots\ldots(21')$$

A similar set of deductions can be made from this table also as were drawn from the previous one. For particular values of ρ_1, ρ_2 and ϕ which a survey would yield, a larger subsample is required to estimate the change when $\rho_1 > \rho_2$ than when $\rho_1 < \rho_2$ whatever the proportion of replacement we might have decided to have. Secondly, the efficiency of the weighted estimate D_W of change relative to the estimate D increases more rapidly when ρ_1 increases than when ρ_2 increases, so that a higher value of correlation between the psu's than that between the corresponding ssu's within them is needed for the efficiency of the estimate D_W of change. Similarly when $\rho_1 > \rho_2$ the efficiency of the estimate would be large for smaller value of S_W^2 as compared to S_b^2 and when $\rho_1 < \rho_2$ a larger value of S_W^2 is needed as compared to S_b^2.

It would be interesting to compare the figures obtained in tables 1 and 2. At q = 1/2 for any values of ρ_1, ρ_2 and ϕ, the change is better estimated than the mean on the second occasion. This reveals two things; firstly the optimum value of q to estimate the change between two occasions is different from the optimum q to estimate the mean on the second occasion, and secondly the optimum value of q to estimate the change would be smaller than the corresponding value to estimate the mean. As a matter of fact $\sigma^2_{D_w}$ attains its minimum value when q = 0 whence its value is then equal to $\frac{2\alpha}{n}(1-\gamma/\alpha)$.

It would be seen from (20) that for any positive value of γ, $\sigma^2_{D_w}$ decreases as γ increases and also q decreases.

Table 2 : Values for the percent efficiency of the estimate D_w of change relative to the estimate D given by (21').

ρ_1	ρ \ ρ_2	m = 3 q = $\frac{1}{2}$						m = 7 q = $\frac{1}{2}$						m = 3 q = $\frac{1}{2}$ ($\frac{1}{3}$)						m=7 q=$\frac{1}{2}$ ($\frac{1}{3}$)					
		.10	.40	.50	.60	.80	.90	.10	.40	.50	.60	.80	.90	.10	.40	.50	.60	.80	.90	.10	.40	.50	.60	.80	.90
	.10	12	13	13	13	14	15	12	13	13	13	13	13	10	11	11	11	12	12	10	11	11	11	11	12
.50	.50	8	12	13	13	16	17	10	12	13	13	14	15	7	10	11	12	14	16	9	10	11	12	13	13
	1.00	7	11	13	15	20	23	9	12	13	13	16	17	6	9	11	14	17	20	7	10	11	12	14	15
	10.00	1	8	13	20	49	86	2	9	13	18	36	53	1	7	11	17	44	75	2	8	11	16	31	46
	.10	36	38	40	41	42	43	38	40	40	41	41	41	32	34	34	36	36	38	35	36	36	36	36	36
.70	.50	24	32	35	38	43	49	32	36	37	38	43	43	20	28	31	33	38	44	29	31	35	36	38	38
	1.0	17	27	32	36	49	56	27	32	35	38	43	46	16	24	27	33	44	50	24	29	31	35	38	43
	10.	2	10	18	25	69	120	3	14	21	28	60	93	2	8	15	21	61	110	4	13	18	25	54	82
	.10	69	74	75	76	80	82	74	76	77	79	80	80	60	65	66	67	71	71	66	67	69	70	71	71
.80	.50	41	53	58	64	80	86	56	64	68	72	80	84	36	47	54	59	71	77	50	58	62	66	71	76
	1.0	27	41	48	56	80	101	43	56	62	68	80	86	24	36	44	50	71	87	38	50	54	62	71	80
	10.0	2	12	21	30	80	161	6	18	27	36	80	130	2	10	15	27	71	139	5	16	22	31	71	116
	.10	145	161	170	180	198	202	170	180	190	200	202	202	130	139	153	160	176	180	156	162	165	170	180	180
.90	.50	77	101	116	131	180	202	118	140	150	161	180	202	65	87	101	125	162	180	105	127	133	143	159	180
	1.0	41	69	85	101	161	202	80	110	121	132	170	202	36	61	71	87	139	180	71	92	107	120	146	180
	10.0	3	14	24	34	93	202	7	24	34	44	110	202	3	12	18	31	76	180	7	21	29	42	99	180

Modification when there are unequal psu's or ssu's.

A situation is very likely to arise when the sample size on the second occasion is different from that adopted on the preceeding one. This may arise due to a number of causes as, for instance, the enquiry conducted on the first occasion may suggest a better estimate for the second occasion if a larger sample be taken on that occasion or extremely unusal weather conditions may necessitate some of the areas to be excluded from the second round of enquiry; such a change in the number of sampling units may bring about some disturbance in the scheme of the survey. It may, therefore, be of interest to study the changes brought about by any such change.

Suppose that out of the n_1 psu's selected on the first occasion, n_1' of them are also retained for the second occasion, while the remaining $n_1'' = n_1 - n_1'$ units are not common with the units on the second occasion. Suppose that we select afresh n_2'' psu's $(n_2'' = n_2 - n_2',\ n_2' = n_1' = n')$ on the second occasion.

If the sample estimates on the two occasions be denoted by $\bar{x}_1', \bar{x}_1''$ and $\bar{y}_2', \bar{y}_2''$, keeping under view the units on which these estimates are based, then

$$\left.\begin{aligned} V(\bar{x}_1') &= V(\bar{y}_2') = \frac{\alpha}{n'} \\ V(\bar{x}_1'') &= \frac{\alpha}{n_1''} \quad ; \quad V(\bar{y}_2'') = \frac{\alpha}{n_2''} \end{aligned}\right\} \qquad \ldots.(22)$$

The estimate of the population mean on the second occasion is then given by

$$E_2 = \frac{n'/n_2}{\left(1-\frac{\gamma^2}{\alpha^2}\frac{n''_1 \; n''_2}{n_1 \; n_2}\right)} \left\{ \bar{y}'_2 + \frac{\gamma}{\alpha}\left(\bar{X}_1 - \bar{x}'_1\right) \right\} + \frac{n''_2\left(1-\gamma^2/\alpha^2 \, \frac{n''}{n_1}\right)}{n_2\left(1-\frac{\gamma^2}{\alpha^2}\frac{n''_1 \cdot n''_2}{n_1 \; n_2}\right)} \bar{\bar{y}}''_2 \quad ..(23)$$

and

$$V(E_2) = \frac{n''_2\left(1-\gamma^2/\alpha^2 \, n''_1/n_1\right)}{n_2\left(1-\frac{\gamma^2}{\alpha^2}\frac{n''_1 \; N''_2}{n_1 \; n_2}\right)} V(\bar{y}''_2) \quad \ldots\ldots(24)$$

In this case

$$\psi_2 = \frac{n''_2\left(1-\gamma^2/\alpha^2 \, n''_1/n_1\right)}{n_2\left(1-\frac{\gamma^2}{\alpha^2}\frac{n''_1 \; n''_2}{n_1 \; n_2}\right)} \quad \ldots\ldots\ldots(24')$$

Sampling on h occasions :- The results developed in the preceeding pages may be extended to any number of occasions. Some assumptions are, however, necessary which we shall mention first. The correlations ρ_1 between the psu's and ρ_2 between the ssu's within each psu are assumed to be constants for the units on any two consecutive occasions. The variance components between psu's and those between ssu's within each psu are assumed to be the same from one occasion to the other. The sampling units are assumed to be drawn with replacement on each occasion. A fixed proportion of units, is replaced on each consecutive occasion so that the size of the sample remains unaltered on each occasion.

Let $\bar{y}_h'$, $\bar{y}_h''$ be the estimates per ssu for the hth occasion based on npm and nqm units respectively; and $\bar{x}_{h-1}'$, $\bar{x}_{h-1}''$ being those for the (h-1)th occasion based on the same number of units, the npm units are common to both h and (h-1)th occasions. An efficient estimate for the population mean on the hth occasion is given by

$$E_h = a_h \bar{Y}_{h-1} + b_h \bar{x}_{h-1}' + c_h \bar{y}_h' + d_h \bar{y}_h''$$

where a_h, b_h, c_h and d_h are subject to certain restrictions to be determined and $\bar{Y}_{h-1}$ provides an efficient estimate for the mean on the (h-1)th occasion.

The condition of unbiasedness gives

$$a_h + b_h = 0 \;; \quad c_h + d_h = 1$$

Therefore

$$E_h = a_h (\bar{Y}_{h-1} - \bar{x}_{h-1}') + (1 - d_h) \bar{y}_h' + d_h \bar{y}_h'' \quad \ldots .(25)$$

The restrictions imposed on a_h and d_h are similar to those imposed by Patterson (1950) for single stage units. These may be written as

$$\mathrm{Cov}\,(\bar{x}'_{h-1}, E_h) = \mathrm{Cov}\,(\bar{Y}_{h-1}, E_h) \quad \ldots\ldots(26)$$

$$\mathrm{Cov}\,(\bar{y}'_h, E_h) = \mathrm{Cov}\,(\bar{y}''_h, E_h) \quad \ldots\ldots(27)$$

and

$$\mathrm{Cov}\,(\bar{y}'_h, \bar{Y}_{h-1}) = \gamma/\alpha \; \mathrm{Cov}\,(\bar{x}'_{h-1}, \bar{Y}_{h-1}) \quad \ldots(28)$$

conditions (26) and (27) follow from the efficiency conditions established by Patterson. Conditions (28) follows from the correlation system. This holds for single stage units. That this holds for two-stage units also may be verified by taking a case of two or three occasions.

For $h = 2$ we have

$$\mathrm{Cov}\,(\bar{y}'_2, \bar{Y}_1) = \mathrm{Cov}\,(\bar{y}'_2, p\,\bar{x}'_1 + q\,\bar{x}''_1)$$

$$= \frac{\gamma}{\alpha}\,\frac{\alpha}{n} = \gamma/\alpha \cdot \mathrm{Cov}\,(\bar{x}'_1, \bar{Y}_1)$$

For $h = 3$

$$\mathrm{Cov}\,(\bar{y}'_3, \bar{Y}_2) = \mathrm{Cov}\left\{\bar{y}'_3,\; a_2(\bar{x}''_1 - \bar{x}'_1) + c_2\,\bar{y}'_2 (1-c_2)\,\bar{y}''_2\right\}$$

$$= \gamma/\alpha \;\frac{q\,(1 - \frac{\gamma^2}{\alpha^2} q)\,\alpha}{(1 - \frac{\gamma^2}{\alpha^2} q^2)\; nq}$$

$$= \gamma/\alpha\,(1 - c_2)\,\frac{\alpha}{nq}$$

Now

$$\mathrm{Cov}\,(\bar{y}'_2, \bar{Y}_2) = \mathrm{Cov}\left\{\bar{y}'_2,\; a_2(\bar{x}''_1 - \bar{x}'_1) + c_2\,\bar{y}'_2 + (1-c_2)\,\bar{y}''_2\right\}$$

$$= \frac{\alpha}{nq}\,(1 - c_2)$$

This proves the relation in (28).

Also we know that
$$\left.\begin{aligned} V(\bar{Y}_{h-1}) &= \mathrm{Cov}\,(\bar{x}''_{h-1}, \bar{Y}_{h-1}) \\ &= \mathrm{Cov}\,(\bar{x}'_{h-1}, \bar{Y}_{h-1}) \end{aligned}\right\} \quad \ldots\ldots(29)$$

Therefore (26) gives

$$\mathrm{Cov}\left[\bar{x}'_{h-1}, a_h\,(\bar{Y}_{h-1}-\bar{x}'_{h-1}) + (1-d_h)\,\bar{y}'_h + d_h\,\bar{y}''_h\right]$$

$$=\mathrm{Cov}\left[\bar{Y}_{h-1}, a_h\,(\bar{Y}_{h-1}-\bar{x}'_{h-1}) + (1-d_h)\,\bar{y}'_h + d_h\,\bar{y}''_h\right]$$

$$\text{i.e. } a_h = \gamma/\alpha\,(1-d_h) \quad \ldots\ldots\ldots\ldots \quad (30)$$

Therefore (25) gives

$$\bar{Y}_h = (1-d_h)\left[\bar{y}'_h + \frac{\gamma}{\alpha}\,(\bar{Y}_{h-1}-\bar{x}'_{h-1})\right] + d_h\,\bar{y}''_h \quad \ldots\ldots(31)$$

as the required estimate for the population mean on the hth occasion.

(27) gives on simplification

$$(1-d_h)\left[\frac{\alpha}{np}\left(1-\frac{\gamma^2}{\alpha^2}\right) + \frac{\gamma^2}{\alpha^2}\,V(\bar{Y}_{h-1})\right] = d_h\,\frac{\alpha}{nq} \quad \ldots(32)$$

$$\text{Also } V(\bar{Y}_h) = \mathrm{Cov}\,(\bar{y}''_h, \bar{Y}_h) = d_h\,\frac{\alpha}{nq} \quad \ldots\ldots(33)$$

$$\text{and } V(\bar{Y}_{h-1}) = \mathrm{Cov}\,(\bar{y}''_{h-1}, \bar{Y}_{h-1}) = d_{h-1}\,\frac{\alpha}{nq} \quad \ldots\ldots(34)$$

Therefore (32) gives

$$(1-d_h) = \frac{p}{\left(1-\frac{\gamma^2}{\alpha^2}\,q + \frac{\gamma^2}{\alpha^2}\,p\,d_{h-1}\right)} \quad \ldots\ldots(35)$$

as the required recurrence relationship between d_h and d_{h-1}, a relation similar to one obtained by Patterson.

The results obtained earlier for two occasions can be obtained from this by putting $h = 2$ and noting that $d_1 = q$.

Suppose now that sampling and consequently replacement also, has been carried over a sufficient number of occasions, then writing $d_h = d_{h-1} \equiv d$ when h is sufficiently large, (35) may be written as

$$d^2 r^2 p + d\,(\alpha^2 - r^2) - q\,(\alpha^2 - r^2) = 0 \qquad \ldots\ldots(36)$$

Solving this for d we have

$$d = \frac{-\left(1 - \gamma^2/\alpha^2\right) + \sqrt{\left(1 - \gamma^2/\alpha^2\right)\left\{1 - \gamma^2/\alpha^2\,(1 - 4pq)\right\}}}{2\,p\,\gamma^2/\alpha^2} \qquad \ldots(37)$$

as the possible solution for d.

In order to obtain the optimum number of psu's which should be replaced on each occasion we have simply to solve the equation $\frac{d}{dq}\, V(\bar{Y}_h) = 0$ where $d_h = d$ in Y_h is given by (37) above.

This gives us on simplification

$$(1 - 2q)\left[2q\,\gamma^2/\alpha^2\,(1 - \gamma^2/\alpha^2) - 4\,d\,\frac{\gamma^4}{\alpha^4}\,(1 + d - \gamma^2/\alpha^2)\right] = 0$$

which gives $q = 1/2$. This shows that after sampling has been carried over sufficiently large number of occasions, then on any occasion 50% of the psu's may be retained from the preceeding occasion and 50% selected afresh. At what stage of sampling this replacement is to be adopted, depends upon practical consideration. Possibly after 3 or 4 occasions, $q = 1/2$ may serve as a good approximation to decide about the replacement policy.

Difference of two estimates:- The estimate Y_h can usually be obtained at each stage and the estimate of change between two successive occasions can be obtained by the difference $\bar{Y}_h - \bar{Y}_{h-1}$. The variance of this difference is, therefore, of interest. For this we need to get Cov $(\bar{Y}_h , \bar{Y}_{h-1})$. Therefore

$$\mathrm{Cov}\,(\bar{Y}_h , \bar{Y}_{h-1}) = \mathrm{Cov}\Big[\bar{Y}_{h-1};\ (1-d_h)\Big\{\bar{y}_h' + \frac{\gamma}{\alpha}(\bar{Y}_{h-1} - \bar{x}_{h-1}')\Big\} + d_h\,\bar{y}_h''\Big]$$

$$= (1 - d_h)\, d_{h-1}\ \gamma/nq.$$

Therefore $V(\bar{Y}_h - \bar{Y}_{h-1}) = V(\bar{Y}_h) + V(\bar{Y}_{h-1}) - 2\,\mathrm{Cov}\,(\bar{Y}_h , \bar{Y}_{h-1})$

$$= d_h \frac{\alpha}{nq} + d_{h-1}\frac{\alpha}{nq} - 2\,(1-d_h)\, d_{h-1}\,\frac{\gamma}{nq}$$

$$\doteq 2\,d\,\Big[\,1 - (1-d\,)\,\gamma/\alpha\Big]\,\frac{\alpha}{nq} \qquad \ldots(38)$$

Table 3. gives the efficiency of the estimate $\bar{Y}_h - \bar{Y}_{h-1}$ of change relative to the difference of the overall means on the two occasions. This is given by the quantity

$$\Big(\frac{\sigma_D^2}{V(\bar{Y}_h - \bar{Y}_{h-1})} - 1\Big) \times 100 .$$

Only one value of $q = 1/2$ has been considered. The efficiency when $q = 1/3$ will definitely be less than when $q = 1/2$; this difference increases as ρ_1 or ρ_2 increase. This result which is evident from table 2, is also true for table 3. A comparison with values in table 2 will show that the efficiency of the estimate of change $\bar{Y}_h - \bar{Y}_{h-1}$ relative to the difference of the overall means is less than the corresponding value when only two occasions are taken. This reduction may possibly be due to the use of limiting value of the weight $d_h = d$. It is needless to say that if this efficiency had been computed relative to difference of the means, had independent samples been taken on the h and (h-1)th occasions, then its value would have been larger than the values obtained above.

In obtaining the estimate of change and its variance we could, as well, have utilised the information provided by the sample on the hth occasion and obtained a modified estimate of the mean on the (h-1)th occasion. Naturally the efficiency of an estimate of change so obtained would be increased. The case has already been dealt with for two occasions only and it can be extended to h occasions also. We denote this modified estimate for the mean on the (h-1)th occasion by ${}_h\bar{Y}_{h-1}$.

Following Patterson, an efficient estimate for the mean on the (h-1)th occasion is given by

$$E_{h-1} = \bar{Y}_{h-1} - w\,\bar{Y}_h + w\,\bar{y}_h''$$

w being the weight to be determined.

$$\frac{d}{dw} V\,(E_{h-1}) = 0 \text{ gives } \quad w = d_{h-1}\,\gamma/\alpha \qquad \ldots.(39)$$

Therefore $\quad {}_h\bar{Y}_{h-1} = \bar{Y}_{h-1} - d_{h-1}\,\gamma/\alpha\;\bar{Y}_h + d_{h-1}\,\gamma/\alpha\;\bar{y}_h'' \quad ..(40)$

Therefore

$$\bar{Y}_h - {}_h\bar{Y}_{h-1} = \bar{Y}_h\,(1 + \gamma/\alpha\; d_{h-1}) - \bar{Y}_{h-1} - d_{h-1}\,\gamma/\alpha\;\bar{y}_h'' \quad ..(41)$$

$$\mathrm{Cov}\,(\bar{Y}_h\,,\,{}_h\bar{Y}_{h-1}) = \mathrm{Cov}(\bar{Y}_h\,,\bar{Y}_{h-1} - d_{h-1}\,\gamma/\alpha\,\bar{Y}_h + d_{h-1}\,\gamma/\alpha\,\bar{y}_h'')$$

$$= \text{Cov}\left(\bar{Y}_h \;,\; \bar{Y}_{h-1} \right)$$

$$= (1 - d_h)\, d_{h-1} \frac{\Upsilon}{nq} \qquad \ldots\ldots(42)$$

Now

$$V({}_h\bar{Y}_{h-1}) = \text{Cov}\left({}_h\bar{Y}_{h-1} \,, \bar{Y}_{h-1} - d_{h-1} \Upsilon/\alpha\, \bar{y}_h + d_{h-1} \Upsilon/\alpha\, \bar{y}_h'' \right)$$

$$= d_{h-1} \frac{\alpha}{nq} - \Upsilon^2/\alpha\; d_{h-1}^2 \frac{(1 - d_h)}{nq} \qquad ..(43)$$

Therefore

$$V(\bar{Y}_h - {}_h\bar{Y}_{h-1}) = d_h \frac{\alpha}{nq} + \left\{ d_{h-1} \frac{\alpha}{nq} - \frac{\Upsilon^2}{\alpha} \frac{d_{h-1}^2 (1-d_h)}{nq} - 2(1-d_h)\, d_{h-1} \frac{\Upsilon}{nq} \right\}.$$

$$\simeq 2\, d \frac{\alpha}{nq} \left[1 - (1-d) \frac{\Upsilon}{\alpha} \right] - \Upsilon^2/\alpha^2\, d^2 (1-d) \frac{\alpha}{nq} \qquad ...(44)$$

Comparing (44) with (38), $V(\bar{Y}_h - \bar{Y}_{h-1})$ is greater than $V(\bar{Y}_h - {}_h\bar{Y}_{h-1})$ the difference in them being equal to $\frac{\Upsilon^2}{\alpha^2} d^2 (1-d) \frac{\alpha}{nq}$, showing that the estimate $(\bar{Y}_h - {}_h\bar{Y}_{h-1})$ is more efficient than $(\bar{Y}_h - \bar{Y}_{h-1})$.

Table: 3

ρ_1	$\phi \backslash \rho_2$	m = 3, q = 1/2: .10	.40	.50	.60	.80	.90	m = 7, q = 1/2: .10	.40	.50	.60	.80	.90
.50	.10	11	12	12	12	13	13	11	12	12	12	12	13
	.50	8	11	12	13	16	17	9	11	12	13	16	17
	1.0	6	10	12	14	17	20	8	11	12	13	17	18
	10.0	1	7	12	16	40	75	2	7	12	17	27	40
.70	.10	27	30	31	32	33	36	31	32	32	32	32	32
	.50	20	25	29	31	36	39	27	31	32	32	36	36
	1.0	17	20	25	29	36	42	23	27	31	32	36	42
	10.0	2	9	15	21	54	81	5	16	18	22	49	70
.80	.10	51	55	56	58	60	61	55	56	58	60	60	61
	.50	31	40	46	51	60	66	38	46	54	58	60	66
	1.0	23	31	37	42	60	72	36	38	49	54	60	68
	10.0	2	9	17	26	60	100	6	18	20	27	60	90
.90	.10	101	119	122	130	180	182	122	130	132	126	182	182
	.50	55	72	80	90	130	182	82	96	98	119	130	182
	1.0	31	54	60	72	119	182	60	80	82	90	130	182
	10.0	3	16	20	29	70	182	8	20	27	36	80	182

Sampling on two occasions with replacement among ssu's only.

Suppose that all the psu's are retained on the second occasion but in each psu only some of the secondstage units are retained on the second occasion while the rest are taken afresh from each selected psu. Let p be the proportion of ssu's retained and q = 1 - p, the proportion replaced, a uniform replacement being adopted for each psu. Then denoting the sample means based on nmp and nmq units on the first and the second occasions by $\bar{x}'$, $\bar{x}''$ and $\bar{y}'$, $\bar{y}''$ respectively, we shall have, following the assumptions made earlier,

$$V(\bar{x}') = \frac{s_b^2}{n} + \frac{s_w^2}{nmp}$$

$$= V(\bar{y}')$$

$$\text{Cov}(\bar{x}', \bar{y}') = \frac{\rho_1 s_b^2}{n} + \frac{\rho_2 s_w^2}{nmp}$$

Since the same psu's are taken on both the occasions, we shall have

$$\text{Cov}(\bar{x}', \bar{y}'') = \frac{\rho_1 s_b^2}{n}$$

$$= \text{Cov}(\bar{x}'', \bar{y}'')$$

$$\text{Cov}(\bar{x}', \bar{x}'') = \frac{s_b^2}{n}$$

$$= \text{Cov}(\bar{y}', \bar{y}'')$$

and $$\text{Cov}(\bar{x}'', \bar{y}') = \frac{\rho_1 s_b^2}{n}$$

$$= \text{Cov}(\bar{x}'', \bar{y}'')$$

The estimate of the population mean on the second occasion is given by

$$E_2 = a\,(\bar{x}'' - \bar{x}') + c\,\bar{y}' + (1 - c)\,\bar{y}''$$

where a and c are obtained by minimising the variance of E_2; this gives us

$$a = \frac{p\,q\,\rho_2}{(1 - \rho_2^2 q^2)} \quad \text{and} \quad c = \frac{p}{(1 - \rho_2^2 q^2)}$$

Therefore

$$E_2 = \frac{p}{(1 - q^2)} \left[\bar{y}' + \rho_2 (\bar{X} - \bar{x}') \right] + \frac{q\,(1 - \rho_2^2 q)}{(1 - \rho_2^2 q^2)}\,\bar{y}''$$

and

$$V(E_2) = \frac{S_b^2}{n} + \frac{q(1 - \rho_2^2 q)\; S_w^2}{(1 - \rho_2^2 q^2)\; nmq}$$

The estimate and its variance are thus seen to be independent of the correlation ρ_1 between the psu's.

Since the total number of units is the same in this as well as the previous case, it is possible to make a comparative study of the two cases. Thus the estimate E_2 will be more, or equal or less efficient than the estimate $\bar{Y}_2$ according as

$$V(E_2) \lesseqgtr V(\bar{Y}_2)$$

i.e. according as

$$\frac{S_b^2}{n} + \frac{q(1 - \rho_2^2 q)\, S_w^2}{(1 - \rho_2^2 q^2)nmq} \lesseqgtr \frac{(m+\phi)\, S_b^2}{m\,n} \cdot \frac{(m+\phi)^2 - q(\rho_1 m + \rho_2 \phi)^2}{(m+\phi)^2 - q^2 (\rho_1 m + \rho_2 \phi)^2}$$

i.e. according as

$$m\, \frac{(\rho_1 m + \rho_2 \phi)^2}{(m+\phi)^2}\,(1 - \rho_2^2 q^2) \lesseqgtr \phi \left[\rho_2^2 - \frac{(\rho_1 m + \rho_2 \phi)^2}{(m+\phi)^2} \right].$$

Since the left hand side is always positive, E_2 will be more efficient than $\bar{Y}_2$ if $\rho_2 > \rho_1$

A study of this case is of particular interest when the costs involved in a survey are also taken into consideration; and especially when a heavy work is expected at the psu level; say, the introduction of a new tehsil in the sample may require at least a week to be spent in preparing the basic frame. In that case by repeating the enquiry on the same set of psu's we sake a lot in terms of money and time which would not have been possible had different psu's been taken on each occasion.

Estimate of change:- An estimate of change may be seen to be the weighted estimate of the quantities $(\bar{y}' - \bar{x}')$ and $(\bar{y}'' - \bar{x}'')$, the weights being $\frac{p}{(1-\rho_2 q)}$ and $\frac{q(1-\rho_2)}{(1-\rho_2 q)}$

i.e. change = $\frac{p}{(1-\rho_2 q)}(\bar{y}' - \bar{x}') + \frac{q(1-\rho_2)}{(1-\rho_2 q)}(\bar{y}'' - \bar{x}'')$

and its variance is given by

$$V(\text{change}) = 2(1-\rho_1)\frac{s_b^2}{n} + \frac{2(1-\rho_2)s_b^2}{(1-\rho_2 q)\,n\,m}$$

Thus for this estimate of change to be efficient, ρ_1 should be large and q should be very small.

Similar results for estimating the mean and the change between any two occasions can easily be obtained, when any number of occasions are taken into consideration.

A COST FUNCTION FOR THE SCHEME FOR THE STUDY OF THE MILK YIELD OF BOVINES IN THE PUNJAB STATE (1956-57)

*****

The field work and the layout of the scheme has been given earlier. Although a three-stage sampling plan was adopted for the scheme, with tehsils, villages and households within villages constituting the three-stages, however, to set up a cost function for the scheme, only two stages have been considered viz. the tehsils and the villages. Actually these are the two stages involving bulk of the expenditure.

The cost function for the summer season of the survey may then be written as

$$C' = c_0' n^{\frac{1}{2}} + c_1' n + c_2' n m^{\frac{1}{2}} + c_3' n m \qquad \ldots\ldots\ldots (45)$$

where C', c_0', c_1', c_2' and c_3' are defined as follows:-

C' = total cost of the survey in the particular season minus the overhead expenditure in that season on stationery, contingencies and maintenance of statistical staff not directly related to the field work.

c_0' = cost of travelling from one tehsil to the other for enumeration or supervision purposes.

c_1' = cost of stay at the tehsil for preparing the basic frame for the selection of the sample etc.

c_2' = cost of travelling from one village to the other within the same tehsil or from the tehsil to the selected village and back assuming that the same amount of travelling would be involved in travelling between two villages within the same tehsil or from tehsil to the selected village.

c_3' =cost of preparing the list of households and selecting a sample thereof for collecting the requisite information.

As mentioned earlier let us assume that only a fraction p of the tehsils selected in the summer season has been retained for re-enumeration in the rainy season and a fraction q = 1-p of them has been selected afresh in the rainy season. In that case there is no need to stay in the tehsil which has been retained, because the frame and list of villages has already been prepared; only thing is that a fresh sample of villages or towns shall have to be drawn. But for the tehsils which have been selected afresh the same process of preparing the requisite frame shall have to be repeated and thus in (45) above $c_1' n$ will be changed to $c_1' n q$; other components of cost will remain unaltered. The cost function for the rainy season would then be

$$C'' = c_o'' n^{\frac{1}{2}} + c_1'' n q + c_2'' n m^{\frac{1}{2}} + c_3'' n m \qquad \ldots (46')$$

Where C'', c_o'' etc. may or may not be different from the corresponding values in (45).

Assuming that total cost incurred on the survey is equally spread out in the three seasons, we shall write $C = C' = C''$, $c_o = c_o' = c_o''$ etc. and the cost function for the two seasons combined together is given by

$$C = c_o n^{\frac{1}{2}} + \frac{1}{2} c_1 (1+q) n + c_2 n m^{\frac{1}{2}} + c_3 n m \quad \ldots\ldots (46)$$

We shall now compute the values for the cost coefficients for the particular scheme under study. For this we must take into consideration the travelling and the daily allowances, which the field workers are entitled to draw when on duty which are needed to compute the expenditure per mile of travel or per unit of time.

The total expenditure on the scheme was about Rs.1.03 lakhs, out of which the overhead expenditure was about Rs.40000/-. Therefore the total expenditure minus the overhead expenditure for the entire scheme comes to about Rs. 63,000/- or roughly Rs. 21,000/- per season. To compute the other components of cost we must chalk out the probable tour itinerary of the field personnel. Travel from one tehsil to another is expected to take about half a day and an equal time will be taken in going from one village to another within the same tehsil, for the reason that while metalled roads are generally available for going from one tehsil to another, one can, at best, get a horse or a bullock cart for travel from one village to another. A field officer has to ~~go to~~ all the selected tehsils. In the first round of the tehsils he will spend about a day in each tehsil to give instructions to the junior field staff and to inspect the frame for the detailed field work in each tehsil. He may also visit at least a village in each tehsil to see for himself whether his instructions are being fully carried out. Thus one day's stay in each selected tehsil and a day's stay in each village he chooses to visit, will take him about 2¼ months, since his month consists of about 20 days of field work, and he has to travel from tehsil to tehsil and from tehsil to village and back. In a subsequent round he can visit a second village in each tehsil (without staying in the tehsil this time) and stay in the village for a day or so; the remaining $1\frac{3}{4}$ months will be taken in this process.

There being three inspectors, an inspector has only 5 tehsils under his supervision. He is expected to tour more extensively and visit each selected village in a tehsil at least twice in a season. In the first round he will also stay in each tehsil for a day to get instructions from the field officer. In the subsequent round there is no need to stay there. Thus he may travel eight to ten times to each tehsil to visit every selected village therein and at each visit to stay there for a day or so.

A supervisor's job is to help and render assistance to the enumerator in his work. Since he has to assist all the five enumerators in his field of enquiry, he will spend about $3\frac{1}{2}$ weeks in each tehsil either visiting all the villages therein and spending a day or two in each village or visit only some of the villages and spend a longer time there.

An enumerator has one tehsil assigned to him for his work. In each round consisting of a month he will visit all the four villages spending about a week there. In each month of the season he will visit each village once.

The contributions to c_0 and c_2 by each member of the field staff would be obtained by calculating their respective rates of expenditure per mile of travel separately for travel between the tehsils and for travel within each tehsil. Similarly their contributions to c_1 and c_3 would be obtained by keeping in view the haltage of each member of the field staff in a tehsil or in a village and the daily allowance to which he is entitled. In any case whether a field worker is travelling or working in the field he would be getting his usual pay as when he is travelling, he is still serving to some end. Combining their respective contributions the values of c_0, c_1, c_2 and c_3 work out to be roughly as Rs. 700, 35.25, 243.25 and 169.20 respectively. Contributions to c_1 would be small since the stay in any tehsil would be of one day only in the first round, there being no stay in the subsequent rounds.

Expression (46) may then be written as

$$21,000 = 700\, n^{\frac{1}{2}} + 1/2 \cdot 35.25\,(1+q)n + 243.25\, n\, m^{\frac{1}{3}} + 169.20\, n\, m$$

or

$$595 = 20\, n^{\frac{1}{2}} + 1/2\,(1+q)\, n + 6.9\, n\, m^{\frac{1}{3}} + 4.8\, n\, m \quad \ldots (47)$$

Henceforth we shall take (47) as the cost function to correspond to the general expression given by (46), where now $c_1 = 1$. This is roughly the cost function for the survey. The actual field work might have been different from the one delineated above but still it will give a good approximation to the actual cost function.

The variance expression for the estimate on the second occasion is given by (12) viz.

$$V(\bar{Y}_2) = \frac{\alpha\,(\alpha^2 - r^2 q)}{n\,(\alpha^2 - r^2 q^2)}$$

$$= S_b^2 \; \frac{(m+\emptyset)}{m\,n} \quad \frac{\left[(m+\emptyset)^2 - (\rho_1 m + \rho_2 \emptyset)^2 q\right]}{\left[(m+\emptyset)^2 - (\rho_1 m + \rho_2 \emptyset)^2 q^2\right]} \quad \ldots\ldots(48)$$

The objective is to find the optimum values for n, q and m which, satisfying the cost function (47) would minimise the variance given by (48).

Consider now a function F of the variance and the cost function, given by

$$F \equiv V(\bar{Y}_2) - \lambda\left[c_o\, n^{\frac{1}{2}} + 1/2\, c_1(1+q)\, n + c_2 n\, m^{\frac{1}{2}} + c_3\, n\, m - C\right]$$

where now C, c_o, c_1, c_2 and c_3 are given by (47), c_1 being equal to 1.

Derivatives of F w.r.t. n, m and q when equated to zero give

$$-\frac{(m+\emptyset)\left[(m+\emptyset)^2 - (\rho_1 m+\rho_2\emptyset)^2 q\right]}{m\,n^2\left[(m+\emptyset)^2 - (\rho_1 m+\rho_2\emptyset)^2 q^2\right]}\cdot S_b^2 = \lambda\left[\tfrac{1}{2}\, c_o\, n^{-\frac{1}{2}} + \tfrac{1}{2}\, c_1(1+q) + c_2\, m^{\frac{1}{2}} + c_3\, m\right] \quad \ldots(49)$$

$$\frac{S_b^2}{n\left\{(m+\emptyset)^2 - (\rho_1 m+\rho_2\emptyset)^2 q^2\right\}^2}\left[\left\{(m+\emptyset)^2 - (\rho_1 m+\rho_2\emptyset)^2 q\right\}\left\{\left[(m+\emptyset)^2 - (\rho_1 m+\rho_2\emptyset)^2 q\right]\left(\frac{-\emptyset}{m^2}\right) + \left(1+\frac{\emptyset}{m}\right) \times \left[2(m+\emptyset) - 2\rho_1(\rho_1 m+\rho_2\emptyset)\, q\right]\right\}\right.$$

$$\left. - \left(1+\frac{\emptyset}{m}\right)\left\{(m+\emptyset)^2 - (\rho_1 m+\rho_2\emptyset)^2 q\right\}\left\{2(m+\emptyset) - 2\rho_1 q^2(\rho_1 m+\rho_2\emptyset)\right\}\right]$$

$$= \lambda n\left(\tfrac{1}{2}\, c_2\, m^{-\frac{1}{2}} + c_3\right) \qquad \ldots\ldots(50)$$

and

$$-\frac{S_b^2 (m+\emptyset)\;(\rho_1 m+\rho_2\emptyset)^2\left[(1-2q)(m+\emptyset)^2 + (\rho_1 m+\rho_2\emptyset)^2 q^2\right]}{mn\left[(m+\emptyset)^2 - (\rho_1 m+\rho_2\emptyset)^2 q^2\right]^2} = c_1 n \quad \ldots(51)$$

$$c_o\, n^{\frac{1}{2}} + \tfrac{1}{2}\, c_1(1+q)\, n + c_2\, n\, m^{\frac{1}{2}} + c_3\, n\, m = C \qquad \ldots(52)$$

To solve these equations for the unknown quantities is not very easy job, the final equation in m after eliminating n and q becomes of very high order.

A possible solution, then is to fix the value of q and solve equations (49), (50) and (52) for m and n that is to obtain the sample size for a fixed value of the replacement fraction.

The effect of replacement on the sample size to be taken in the rainy season may be seen by taking a set of values for q.

The value for n will be obtained by solving (52) as a quadratic in n.

Equations (49) and (50) then give

$$\left\{ \tfrac{1}{2} c_0 n^{-\frac{1}{2}} + \tfrac{1}{2} c (1+q) + c\, m^{\frac{1}{2}} + c\, m \right\} \left\{ \left\{ (m+\phi)^2 - (\rho_1 m + \rho_2 \phi)^2 q^2 \right\} \times \right.$$

$$\left[\left\{ (m+\phi)^2 - q(\rho_1 m + \rho_2 \phi)^2 \right\} \left\{ \left(-\frac{\phi}{m}\right) \right\} + 2(m+\phi) \left\{ (m+\phi) - \rho_1 q (\rho_1 m + \rho_2 \phi) \right\} \right] -$$

$$\left. 2(m+\phi) \left\{ (m+\phi)^2 - (\rho_1 m + \rho_2 \phi)^2 q \right\} \left\{ (m+\phi) - \rho_1 q^2 (\rho_1 m + \rho_2 \phi) \right\} \right\} +$$

$$\left(\tfrac{1}{2} c_2 m^{\frac{1}{2}} + c_3 m\right)\left(\frac{m+\phi}{m}\right) \left[(m+\phi)^2 - (\rho_1 m + \rho_2 \phi)^2 q \right] \left[(m+\phi)^2 - (\rho_1 m + \rho_2 \phi)^2 q^2 \right]$$

$$= 0 \quad \ldots (53)$$

If we fix the value of q at $\frac{1}{2}$, then we shall have on simplification

$$\left(\tfrac{1}{2} c_0 n^{-\frac{1}{2}} + .75 + c_2 m^{\frac{1}{2}} + c_3 m\right) \left[\left\{ 1 - \tfrac{1}{2}\left(\frac{\rho_1 m + \rho_2 \phi}{m+\phi}\right)^2 \right\} \left\{ 1 - \tfrac{1}{2}\left(\frac{\rho_1 m + \rho_2 \phi}{m+\phi}\right)^2 \right\} \right.$$

$$\left. \left(\frac{-\phi}{m}\right) + \tfrac{1}{2}\left(\frac{\rho_1 m + \rho_2 \phi}{m+\phi}\right)^2 - \tfrac{1}{2}\left(\frac{\rho_1 m + \rho_2 \phi}{m+\phi}\right) \right] + \left(\tfrac{1}{2} c_2 m^{\frac{1}{2}} + c_3 m\right).$$

$$\left(\frac{m+\phi}{m}\right) \left[1 - \tfrac{1}{2}\left(\frac{\rho_1 m + \rho_2 \phi}{m+\phi}\right)^2 \right] \left[1 - \tfrac{1}{4}\left(\frac{\rho_1 m + \rho_2 \phi}{m+\phi}\right)^2 \right] = 0 \quad \ldots (54)$$

Similarly for $q = \frac{3}{4}$ we shall have

$$\left[\left(\tfrac{1}{2}c_0\, n^{-\frac{1}{2}} + .875 + c_2\, m^{\frac{1}{2}} + c_3 m\right)\right]\left[\left\{1 - \frac{9}{16}\left(\frac{m+\ \ \phi}{m+\phi}\right)^2\right\}\left\{1 - \tfrac{3}{4}\left(\frac{\rho_1 m + \rho_2\phi}{m+\phi}\right)^2\right\}.\right.$$

$$\left.\left(\frac{-\phi}{m}\right) + \tfrac{3}{8}\left(\frac{\rho_1 m + \rho_2\phi}{m+\phi}\right)^2 - \tfrac{3}{8}\rho_1\left(\frac{\rho_1 m + \rho_2\phi}{m+\phi}\right)\right] + \left(\tfrac{1}{2}c_2 m^{\frac{1}{2}} + c_3 m\right) .$$

$$\left(\frac{m+\phi}{m}\right)\left\{1 - \tfrac{3}{4}\left(\frac{m+\ \ \phi}{m+\phi}\right)^2\right\}\left\{1 - \frac{9}{16}\left(\frac{\rho_1 m + \rho_2\phi}{m+\phi}\right)^2\right\} = 0 \quad ..(55)$$

and a similar expression for $q = \frac{1}{2}$ also.

While with the algebraic method it may be difficult to get a solution in m for the above equations for known values of ρ_1, ρ_2, ϕ and cost constants, the method of trial and error shall have to be resorted to. Thus only an approximate value for m is possible which may satisfy the given equations.

Similarly it is also possible to find the optimum values for m and n which would best estimate the change in milk yield of animals from one season to the other, for the same replacement fractions viz. $q = \frac{1}{2}$, $\frac{3}{4}$ and $\frac{1}{4}$ and satisfying the above cost function. Again the method of trial and error shall be used to arrive at a solution.

The equation to be solved for m after minimising (20) w.r.t. m and n is given by

$$\left(\tfrac{1}{2}c_2\, m^{\frac{1}{2}} + c_3 m\right)\left[1 - \left(\frac{\rho_1 m + \rho_2\phi}{m+\phi}\right)\right]\left[1 - \left(\frac{\rho_1 m + \rho_2\phi}{m+\phi}\right) q\right](m+\phi)$$

$$- \left[\tfrac{1}{2}c_0\, n^{-\frac{1}{2}} + \tfrac{1}{2}(1+q) + c_2\, m^{\frac{1}{2}} + c_3 m\right]\left[\rho_1 m(1-q) + q\, m\left(\frac{\rho_1 m + \rho_2\phi}{m+\phi}\right) + \right.$$

$$\left.\phi\left\{1 - q\left(\frac{\rho_1 m + \rho_2\phi}{m+\phi}\right)\right\}\left\{1 - \left(\frac{\rho_1 m + \rho_2\phi}{m+\phi}\right)\right\} - m\left(\frac{\rho_1 m + \rho_2\phi}{m+\phi}\right)\right] = 0 \quad ..(56)$$

where n is again obtained by solving (47) as a quadratic in n for different values of q.

Three equations would be obtained for $q = \frac{1}{2}, \frac{2}{3}$ and $\frac{1}{3}$ each of which is to be solved independently for m.

For the scheme conducted in Punjab the analysis of the data collected on the milk yield of cows yields the following estimates for the population variances and correlation coefficients:-

Estimates of true variance		Summer	Rainy	Winter
Between Tehsils	Est. S_b^2	19.63	30.06	4.52
Between villages within Tehsils.	Est. S_w^2	68.85	58.37	66.17

and correlation coefficients between the three seasons are:-

Summer-rainy	Rainy-winter	Winter-summer
0.6072	0.7298	-0.0787

Considering the values for the summer and rainy seasons only we shall have the pooled mean square for the two seasons as:-

$$\text{Est. } \bar{S}_b^2 = 24.846 \qquad \text{Est. } \bar{S}_w^2 = 63.610$$

Therefore $\phi = 2.56$. and $\rho_1 = 0.6072$; ρ_2 will not occur in the variance formula since in each of the three seasons different villages have been taken for enumeration purpose from each tehsil. Thus to estimate the milk yield of cows in the rainy season, we need to have for different values of the replacement fraction q, the following optimum sizes for m and n which should be taken for the sample.

For $q = 0$ $\quad m = 1.5 \quad n = 29.80$

For $q = \frac{1}{3}$ $\quad m = 1.6 \quad n = 28.10$

For $q = \frac{1}{2}$ $\quad m = 1.7 \quad n = 27.30$

For $q = \frac{2}{3}$ $\quad m = 1.8 \quad n = 27.20$

Similarly for estimating the change in the milk yield of cows from one season to the other, the corresponding values for m and n are:-

For $q = 0$ $\quad m = 2.0 \quad n = 24.80$

For $q = \frac{1}{3}$ $\quad m = 2.2 \quad n = 23.50$

For $q = \frac{1}{2}$ $\quad m = 2.3 \quad n = 23.30$

For $q = \frac{2}{3}$ $\quad m = 2.4 \quad n = 21.50$

While the above are strictly the optimum values for m and n for estimating the mean and the change for the milk yield of cows alone, different sets of values for m and n would be obtained for estimating the average yield and the change from one season to the other in case of buffaloes or for estimating the values for some other statistic for the same data, as the enquiry conducted in Punjab was broad in its scope in as much as, besides estimating the milk yield of bovines, information was also collected on various practices relating to rearing and feeding of animals. The optimum value for the sample size is therefore, likely to differ for estimating the values for different characters. To decide what sample size should be taken is really a question of ascertaining what the reliability of different sample estimates will be and seeing in which character of the survey a loss can be tolerated.

Table No.4 gives the values for the sample size to

Table 4. Values of opt. m and opt. n for $q = \frac{1}{2}$ for a set of values of ρ_1, ρ_2 and $\phi = S_w^2/S_b^2$, satisfying the cost function given by equation (47) and the expression for the variance is given by equation (48).

ρ_2		0.00		0.50		0.70		0.90	
$\rho_1 \downarrow$	ϕ	m	n	m	n	m	n	m	n
.50	2	1.4	30.90	1.3	32.6	1.3	32.6	1.2	34.4
	3	1.8	26.4	1.7	27.3	1.7	27.3	1.6	28.4
	4	2.2	23.3	2.1	24.0	2.1	24.0	1.9	25.6
	5	2.6	20.9	2.4	21.9	2.4	21.9	2.3	22.8
	10	4.0	15.3	3.9	15.6	3.8	15.9	3.7	16.3
	15	5.6	12.0	5.3	12.8	5.2	13.0	5.0	13.0
.70	2	1.5	29.7	1.4	30.9	1.3	32.6	1.2	34.4
	3	1.9	25.6	1.8	26.4	1.7	27.3	1.7	27.3
	4	2.3	22.8	2.2	23.3	2.1	24.0	2.0	24.7
	5	2.7	20.0	2.5	21.3	2.5	21.3	2.4	21.9
	10	4.2	14.7	4.0	15.3	3.9	15.6	3.8	15.9
	15	5.8	11.7	5.5	12.1	5.3	12.8	5.1	13.2
.90	2	1.6	28.4	1.5	29.7	1.4	30.9	1.3	32.6
	3	2.0	24.7	1.9	25.6	1.8	26.4	1.8	26.4
	4	2.5	21.3	2.4	21.9	2.3	22.8	2.1	24.0
	5	2.8	19.2	2.7	20.0	2.6	20.9	2.5	21.3
	10	4.4	14.5	4.2	14.7	4.0	15.3	3.9	15.6
	15	6.0	11.4	5.7	11.8	5.4	12.3	5.2	13.0

Table- 5:- Values of Opt. m and Opt. n which would best estimate the change from one occasion to the other, subject to the cost function given by (47), the expression for the variance of change being given by (20).

p_1	p_2	0.00		0.50		0.70		0.90	
	ϕ	m	n	m	n	m	n	m	n
.50	2	1.6	28.4	1.4	30.9	1.2	34.4	.9	38.5
	3	2.2	23.3	1.8	26.4	1.5	29.7	1.0	37.9
	4	2.7	20.0	2.2	23.3	1.9	25.6	1.2	34.4
	5	3.0	17.2	2.5	21.3	2.1	24.0	1.3	32.6
	10	5.0	13.0	4.0	15.3	3.3	17.4	1.8	26.4
	15	7.5	9.6	6.0	11.4	4.5	14.2	2.3	23.3
.70	2	2.0	24.7	1.5	29.7	1.3	32.6	1.1	36.2
	3	2.6	20.9	2.0	24.7	1.8	26.4	1.5	29.7
	4	3.3	17.4	2.6	20.9	2.4	21.9	1.8	26.4
	5	3.9	15.6	3.1	17.0	2.8	19.2	2.0	24.7
	10	6.0	11.4	5.0	13.0	4.3	14.6	3.0	17.2
	15	8.6	8.6	7.0	10.6	6.3	11.0	4.1	13.1
.90	2	2.4	21.9	2.4	21.9	2.2	23.3	1.2	34.4
	3	3.7	16.3	3.6	16.7	3.4	17.2	1.9	25.6
	4	4.8	13.4	4.6	13.9	4.2	14.7	3.0	17.2
	5	5.3	12.8	5.1	13.2	4.8	13.4	3.8	15.9
	10	8.2	9.0	7.9	9.2	7.4	9.7	5.4	12.3
	15	11.0	7.0	10.6	7.3	10.0	7.7	6.0	11.4

be taken on the second occasion for a set of values of ρ_1, ρ_2, and ϕ which would minimise the variance of the estimate for the second occasion, the cost function being given by (47). In a similar way a set of values for m and n may be obtained for different values of ρ_1, ρ_2 and ϕ which would best estimate the change from one occasion to another and satisfying the given cost function. Table No.5 gives some such values.

A comparison of the figures in tables 4 and 5 would reveal that lesser number of psu's are needed to estimate the change from one occasion to the other than to estimate the mean on the second occasion, if $\rho_1 \geqslant \rho_2$ and vice versa if $\rho_1 < \rho_2$. Consequently the size of the subsample would be larger, for estimating the change, than for estimating the mean if $\rho_1 \geqslant \rho_2$ and smaller, if $\rho_1 < \rho_2$. This justifies the observation made earlier that the optimum q to estimate the change is different from the corresponding optimum to estimate the mean on the second occasion. Thus ρ_1 and ρ_2 happen to play a more significant part in determining the sample size required to estimate the change. It may further be observed that q, the replacement fraction, is not so important in changing the sample size; that is, if we increase the value of q, there is not going to be any significant change in the values of n and m unless ρ_1 and ρ_2 are sufficiently large. This is so because the coefficient $c_1 = 1$ of q is very small as compared to the coefficients of other terms. This may be seen from other considerations also. Since the actual field work is done in the villages and the stay in the tehsils is of very short duration, there is not going to be any substantial reduction in the field work whether the same tehsils are taken on all occasions or some of them are replaced in the second season. A reduction in the field work can be expected only when some of the villages are also retained.

Remark 1: The optimum sample size by using a slightly simpler cost function than the one used above.

We have seen that the effect of travel between the ssu's within psu's is reflected premarily by the term $c_2 \, n \, m^{\frac{1}{2}}$. Suppose now that the field work is so arranged that the travel between the ssu's is reduced to a negligible quantity, so that this term is small in its effect and can be neglected. Then our cost function becomes

$$595 = 20 \, n^{\frac{1}{2}} + \tfrac{1}{2} \, (1 + q) \, n + 4.8 \, n \, m \qquad \ldots\ldots(57)$$

The optimum values for m and n for $q = \frac{1}{2}$ may then be obtained which would best estimate the mean yield of the character on the second occasion as also the change from one occasion to the other. Table No.6 gives the values for m and n separately obtained for estimating the mean and the change, for a set of values of ρ_1, ρ_2 and ϕ.

It may be seen that ommission of the term c_2 of cost has resulted in a considerable reduction of the size of the subsample both for estimating the mean on the second occasion as also for estimating the change from first occasion to the second; consequently there has been an increase in the number of psu's to be selected.

Remark 2: **Optimum sample size when there is an increase in the c_1 component of cost.**

It is sometimes possible that the preparation of basic frame and listing and selection of ssu's from each tehsil may require a longer stay at the tehsil. Such a possibility may occur when we have embarked upon an exploratory survey of a material for which no basic frame is available. Suppose that this component of cost takes five times the value already taken for it. The cost function is then given by

$$595 = 20\, n^{\frac{1}{2}} + 5/2\,(1+q)\, n + 6.9\, n\, m^{\frac{1}{2}} + 4.8\, n\, m \quad ..(58)$$

The values of optimum m and n for this cost function are given in table No. 7.

Table 6: Values of Opt.m and Opt.n which would best estimate (a) the mean on the second occasion, and (b) the change in the yield of the character from first occasion to the second, subject to the cost function $20m^{\frac{1}{2}} + \frac{1}{2}(1+q)n + 4.8\,nm = 595$, q being equal to $\frac{1}{2}$.

......

		Mean								Change							
	ρ	0.00		0.50		0.70		0.90		0.00		0.50		0.70		0.90	
P_1	ϕ	m	n	m	n	m	n	m	n	m	n	m	n	m	n	m	n
.50	2	1.0	37.9	1.0	37.9	0.9	38.5	0.9	38.5	1.2	34.4	1.0	37.9	0.9	38.5	0.8	38.5
	3	1.2	34.4	1.2	34.4	1.1	36.2	1.0	37.9	1.4	30.9	1.2	34.4	1.0	37.9	.9	38.5
	4	1.5	29.7	1.4	30.9	1.3	32.6	1.3	32.6	1.6	28.4	1.4	30.9	1.1	36.2	1.0	37.9
	5	1.7	27.3	1.6	28.4	1.5	29.7	1.5	29.7	1.8	26.4	1.5	29.7	1.2	34.4	1.1	36.2
	10	2.3	22.8	2.2	23.3	2.1	24.0	2.0	24.7	2.6	20.9	2.1	24.0	1.8	26.4	1.6	28.4
	15	3.1	17.6	2.9	18.2	2.8	19.2	2.7	20.0	3.4	17.8	2.8	19.2	2.4	21.9	2.1	24.0
.70	2	1.1	36.2	1.0	37.9	.9	38.5	.9	38.5	1.4	30.9	1.2	34.4	1.1	36.2	.9	38.5
	3	1.3	32.6	1.2	34.4	1.1	36.2	1.0	37.9	1.6	28.4	1.4	30.9	1.3	32.6	1.0	37.9
	4	1.6	28.4	1.5	29.7	1.4	30.9	1.4	30.9	1.8	26.4	1.6	28.4	1.5	29.7	1.3	32.6
	5	1.8	26.4	1.7	27.3	1.6	28.4	1.6	28.4	2.1	24.0	1.8	26.4	1.7	27.3	1.5	29.7
	10	2.5	21.3	2.3	22.8	2.2	23.3	2.1	24.0	3.0	17.8	2.7	20.0	2.5	21.3	2.0	24.7
	15	3.4	17.2	3.2	17.5	3.0	17.8	2.9	18.2	4.0	15.3	3.6	16.7	3.3	17.4	2.8	19.2
.90	2	1.1	36.2	1.0	37.9	1.0	37.9	1.0	37.9	1.5	29.7	1.4	30.9	1.2	34.4	1.1	36.2
	3	1.4	30.9	1.3	32.6	1.2	34.4	1.0	37.9	1.8	26.4	1.6	28.4	1.4	30.9	1.2	34.4
	4	1.7	27.3	1.6	28.4	1.5	29.7	1.4	30.9	2.1	24.0	1.9	25.6	1.6	28.4	1.5	29.7
	5	2.0	24.7	1.8	26.4	1.7	27.3	1.6	28.4	2.4	21.9	2.1	24.0	1.8	26.4	1.6	28.4
	10	2.7	20.0	2.5	21.3	2.4	21.9	2.0	23.3	3.4	17.2	3.0	17.8	2.7	20.0	2.5	21.3
	15	3.7	16.3	3.4	17.2	3.2	17.5	3.0	17.2	4.3	15.0	3.9	15.6	3.6	16.7	3.3	17.4

<u>Table 7:</u> Values of Optimum m and n which would best estimate (a) the mean on the second occasion and (b) the change in the yield of the character from first occasion to the second subject to the cost function 20 n + 5/2 (1+q)n + 6.9 n m + 4.8 mm =595, q being equal to $\frac{1}{3}$.

		Mean								Change							
p_1	p_2 / ø	0.00		0.50		0.70		0.90		0.00		0.50		0.70		0.90	
		m	n	m	n	m	n	m	n	m	n	m	n	m	n	m	n
	2	1.8	23.1	1.6	24.6	1.5	25.5	1.4	26.4	2.1	21.1	1.8	23.1	1.5	25.5	1.2	28.6
	3	2.2	20.6	2.0	21.7	1.9	22.4	1.8	23.1	2.6	18.6	2.3	20.0	1.9	22.4	1.5	25.5
.50	4	2.7	18.2	2.5	19.1	2.4	19.5	2.3	20.0	3.2	16.4	2.9	17.4	2.3	20.0	1.8	23.1
	5	3.2	16.4	3.0	17.0	2.9	17.4	2.8	17.7	3.8	17.6	3.4	15.7	2.8	17.7	2.1	21.1
	10	4.7	12.7	4.5	13.1	4.4	13.3	4.2	13.7	5.3	11.7	4.8	12.5	3.7	14.9	2.5	19.[illegible]
	15	6.2	10.5	5.9	10.8	5.8	11.0	5.5	11.4	7.0	9.5	6.5	10.1	4.8	12.5	3.0	17.[illegible]
	2	1.8	23.1	1.7	23.8	1.6	24.6	1.5	25.5	2.6	18.6	2.4	19.5	2.0	21.7	1.5	25.[illegible]
	3	2.3	20.0	2.1	21.1	2.0	21.7	1.9	22.4	3.2	16.4	2.8	17.7	2.3	20.0	1.6	24.[illegible]
	4	2.8	17.7	2.6	18.6	2.5	19.1	2.4	19.5	3.8	14.6	3.3	16.0	2.7	18.2	1.8	23.[illegible]
.70	5	3.3	16.0	3.1	16.7	3.0	17.0	2.9	17.4	4.6	12.9	4.1	13.9	3.2	16.4	2.0	21.[illegible]
	10	4.9	12.3	4.7	12.7	4.6	12.9	4.4	13.3	7.0	9.5	6.1	10.6	4.8	12.5	2.8	17.[illegible]
	15	6.5	10.0	6.3	10.4	6.2	10.5	6.0	10.7	9.4	7.7	7.8	8.6	6.3	10.4	3.6	15.[illegible]
	2	2.0	21.7	1.9	22.4	1.8	23.1	1.6	24.6	3.4	15.7	3.2	16.4	3.0	17.0	2.3	20.0
	3	2.5	~~20.0~~ 19.1	2.4	19.5	2.3	20.0	2.2	20.6	4.4	13.3	4.2	13.7	4.0	14.1	2.9	17.4
.90	4	3.0	17.0	2.9	17.4	2.8	17.7	2.6	18.6	5.8	11.0	5.6	11.2	5.3	11.7	3.7	14.9
	5	3.5	15.4	3.4	15.7	3.3	16.0	3.1	16.7	6.9	9.6	6.6	10.0	6.2	10.5	4.3	13.5
	10	5.2	11.8	5.1	12.0	4.9	12.3	4.7	12.7	9.6	7.5	8.8	8.1	8.2	8.5	5.4	11.5
	15	6.8	9.8	6.6	10.0	6.4	10.2	6.2	10.5	12.0	5.4	10.0	7.4	8.8	8.1	6.5	10.1

SUMMARY

Application of successive sampling with a two-stage sampling design to some of the problems where the survey is repeated at regular intervals has been indicated. Partial replacement of units is advantageous in field operation. Estimate of the population mean and its variance has been obtained (a) for two occasions and (b) for h occasions, the sampling scheme being that a fixed proportion of the psu's taken on the preceding occasion has been replaced on the current occasion, the ssu's within each selected psu's being completely retained. Also an estimate for the change between any two consecutive occasions and its variance has been obtained under the same sampling scheme. Similar expressions for the estimate and the change and their variances have been obtained under a different sampling pattern, viz. that the same psu's are taken on both occasions, but a fraction p of the ssu's within each selected psu is retained on the second occasion, and a fraction q selected afresh. A comparative study of the two sampling patterns has also been made. Data collected from the scheme conducted in Punjab (1956-57) to study the milk-yield of bovines there has been taken and a cost function for the scheme has been obtained. Optimum size of the sample to be taken in the second season to estimate the milk-yield of cows has been obtained. It has been observed that the size of the sample is not affected if some tehsils are replaced on the second season. Optimum values for n and m to be taken on the second occasion to estimate the mean and the change have been tabulated for a set of values of the correlations ρ_1 and ρ_2 and $\phi = S_w^2/S_b^2$.

R E F E R E N C E S

Jessen, R.J.	(1942)	Statistical Investigation of a Sample Survey for Obtaining Farm Facts; Iowa Agr. Exp.Sta. Res. Bull. 304.
Yates, F.	(1949)	Sampling Methods for Censuses and Surveys, Charles Griffin and Co., London.
Patterson, H.D.	(1950)	Sampling on Successive Occasions with Partial Replacement of Units, J.Roy. Stat. Soc., Series B,12.
Tikkiwal, B.D.	(1953)	Theory of Successive Sampling Jour. Ind. Soc. Agr. Stat.,5.
Narain, R.D.	(1953)	On the Recurrence Formula in Sampling on Successive Occasions. Jour. Ind. Soc. Agr. Stat., 5.
Sukhatme, P.V.	(1953)	Sampling Theory of Surveys with Applications. Indian Soc. of Agri. Stat.,New Delhi. The Iowa State College Press, Ames, Iowa, U.S.A.
Tikkiwal, B.D.	(1956)	Some Further Contribution to the Theory of Univariate Sampling on Successive Occasions; Jour. Ind. Soc. Agr. Stat. 8.
Hansen and Hurwitz	(1956)	Sample Survey Methods and Theory. John Wiley Sons, New York.
Singh,D.	(1959)	Paper read at the 12th Annual Meeting of the Indian Society of Agricultural Statistics held at Gwalior.
I.C.A.R.	(1956)	A preliminary report on the Scheme conducted for the study of milk-yields of breeds and feeding and management practices of cattle and buffaloes in Punjab.