SAMPLING ON TWO OCCASIONS
WITH PARTIAL REPLACEMENT OF UNITS

by 1264

NANTIT SUPAMONGKON

B.A., Chulalongkorn University, 1964

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY

Manhattan, Kansas

1970

Approved by:

A. M. Feyerherm

Major Professor

TABLE OF CONTENTS

Introduction

When a population is subject to change over time, a single survey will only yield information about the properties of that population on a given occasion. A Survey must be repeated on several occasions in order to examine the changes of population characteristics and keep all aspects of the original estimated values up-to-date.

When the same population is sampled repeatedly, the opportunity for flexible sample design are increased. For example, on the second occasion, we may have parts of the sample that are matched with the first occasion. Such a method of partial matching has been termed "sampling on successive occasions with partial replacement of units" or "rotation sampling" or "sampling for a time series".

Yates (5) listed several alternatives for periodical re-survey as follow:

(1) A complete census or survey may be repeated in its original form at intervals.

(2) A sample census or survey may be repeated at intervals, a new sample being selected on each occasion without regard to previous sampling. He suggested the term "independent samples" for this sample scheme.

(3) A sample census or survey may be repeated on the same sample. The suggested term for this type of resurvey is "fixed sample".

(4) Part of the sample may be replaced on each occasion, the remainder being retained. The suggested term is "partial replacement".

(5) A resurvey of a subsample of the original sample may be made. In the case of complete census this is equivalent to a resurvey of a

sample of the whole population. The suggested term for this scheme is "subsample".

The procedure of partial replacement of units in successive sampling was initiated by Jessen (1). On two occasions, he computed two estimates on the second occasions as follows:

(1) a sample mean based on the new sample units

(2) a regression estimate based on the sample units observed on both occasions and an overall sample mean obtained on the first occasion.

A linear unbiased estimate of the population mean on the second occasion was obtained by combining the two estimates.

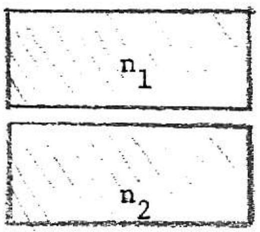
By Yates (5), this procedure was extended from two occasion to n, $(n > 2)$ occasions under the restrictive conditions of constant sample size and fixed replacement rate on each occasion. Patterson (2) generalized the results further. Rao and Graham (3) employed a rotation sample design for these repeated occasion.

This paper is concern with estimate of some population values when alternative (4) with two occasions is adopted.

FIXED SAMPLE, INDEPENDENT SAMPLES AND PARTIAL REPLACEMENT

By comparison, we will clearly see the difference among the methods of fixed sample, independent samples and partial replacement. The advantages of partial replacement over the two schemes will be discussed.

(a) Fixed Sample



First Occasion.

Second Occasion.

Let n_1 = number of sampling units on the first occasion.

n_2 = number of sampling units on the second occasion.

$n_1 = n_2$ and they are the same sample.

Advantages of Fixed Sample

(1) In fixed sample scheme, we are able to evaluate a change of value in a characteristic under study. If there is a difference between values of our estimates from two occasions, we may say that it arises from the change of time.

This scheme is beneficial when used with non-human-populations as in a cropping survey or in forestry research.

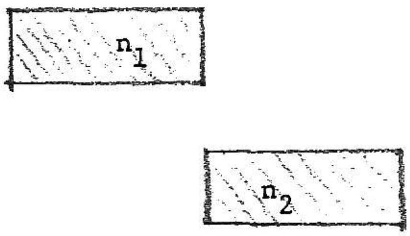
(2) We can reduce cost and save time in survey planning since we use the same sample in the latter occasion.

Disadvantages of Fixed Sample

(1) Difficulty arises when the survey dealing with samples of human-populations. Enumerators frequently find lack of co-operation in giving information from the interviewees on the successive occasions.

(2) In the case that there is a change in our sample boundary such as damage from fire, we cannot substitute the lost sampling units by new selected units. If we do, there will be biasness in our estimate.

(b) Independent Samples



First Occasion.

Second Occasion.

n_1 and n_2 are defined as in a fixed sample scheme, but they are different samples from the same population.

Advantages of Independent Samples

(1) We get better co-operation from interviewees in fieldwork since we use a new sample on every occasion.

(2) Information based on all occasions yields an efficient total estimate. This is equivalent to simply increasing the number of sampling units.

Disadvantages of Independent Samples Scheme

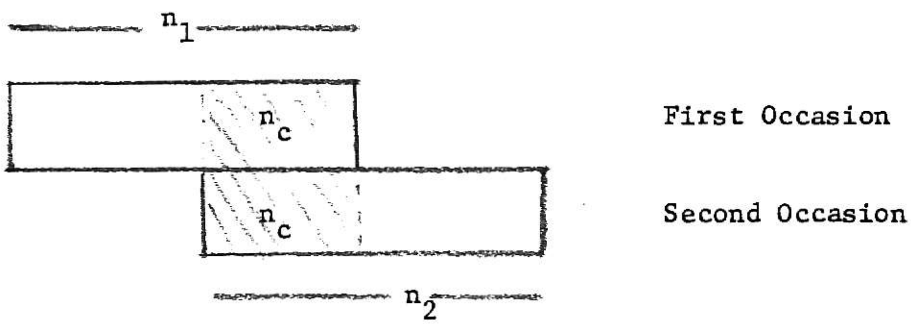
(1) It involves higher costs and more time in preparing a new sample on every occasion.

(2) We cannot exactly know whether the change of estimated values occurred from time series. For example, in the case of sampling from human-populations, the change might arise from difference of behavior such as behavior in expenditure.

It will be noted that the fixed sample scheme is formally equivalent to the observation of different characteristics (variates) on the same sample and independent samples are formally equivalent to interpenetrating sample. (Yates, 1965, p. 46)

The disadvantages of fixed sample and independent samples schemes can be remedied by the scheme of partial replacement. Theoretically, we can show that the estimates from partial replacement is more efficient than the first two schemes.

(c) Partial Replacement



n_1 = number of sampling units on the first occasion.

n_2 = number of sampling units on the second occasion.

n_c = number of repeated sampling units on the two occasions.

Advantages of Partial Replacement

(1) We can reduce the problem of lack of co-operation from human samples in repeating our survey on the same units. The trouble in this case will be at most from n_c units.

(2) Cost of survey planning is less than independent samples.

(3) We are able to evaluate estimated changes, estimated mean and estimated variance at the same time. The change of values can be estimated because parts of the sample are repeated. The mean for the last occasion, mean for all occasions and total for all occasions are also estimated.

(4) In the case there is a change in a sample boundary, replacement by new sampling units for the lost ones can be done without biasness.

REPLACEMENT RATE

The number of repeated units n_c depends upon the replacement rate. If we let U equal the replacement rate, $0 < U < 1$, we can see that when $U = 0$, the sample scheme is fixed sample and when $U = 1$, the sample scheme is independent samples. We may specify $U = \frac{1}{4}, \frac{1}{2}$ or possibly a number which yields optimum results.

Example 1.

Suppose we want to draw sample of size 8 units from a certain population,

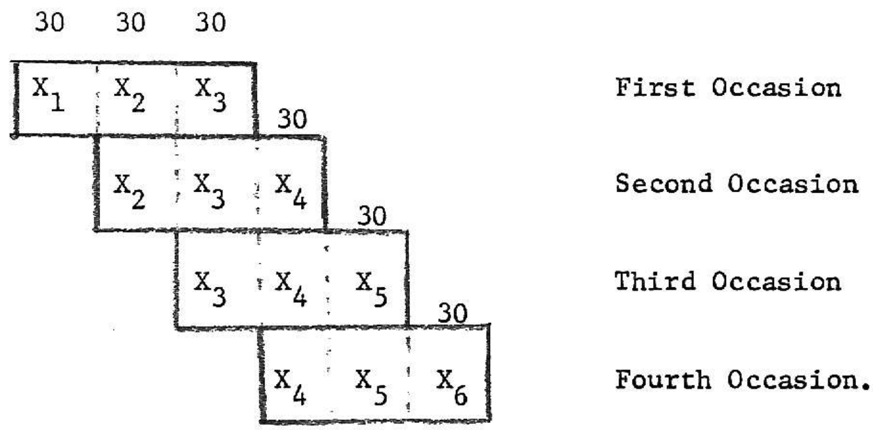
$\boxed{X_1, X_2 \mid X_3, \ldots\ldots, X_8}$ First Occasion

$\boxed{X_3, \ldots\ldots, X_8 \mid X_9, X_{10}}$ Second Occasion

The measured values X_3, X_4, ..., X_8 are the repeated sampling units. Rate of replacement $U = \frac{2}{8} = \frac{1}{4}$. The number of new selected units $= 2 = Un$, and the number of repeated units $= 6 = (1 - U)n$.

Example 2.

In the case of four successive occasions,



Let n = number of sampling units on each occasion = 90 units

p = number of occasions = 4.

U = rate of replacement $= \frac{1}{3}$.

Then,

Un = the number of new selected units on each successive occasion = 30 units.

$(1 - U)n$ = number of units repeated in each successive sample = 60 units.

$n+(p - 1)Un$ = total number of distinct units selected = 180 units.

METHOD OF ESTIMATION

In sampling theory, we are interested in the procedures of sampling designs and estimation of parameters based on information from these designs.

General Procedure for Estimation

(a) Mean

In the case of single-occasion unequal probability sampling, the population mean will be in the form of

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{P_i \; nN}$$

where N = total number of units in the population

n = number of sampling units

P_i = arbitrary probability, $0 < P_i < 1,$

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

We see that when $P_i = \frac{1}{N}$ for all i, the sample is selected with equal probability. The mean value will be in the form

$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$, which is the result for a simple random sampling design.

(b) Total

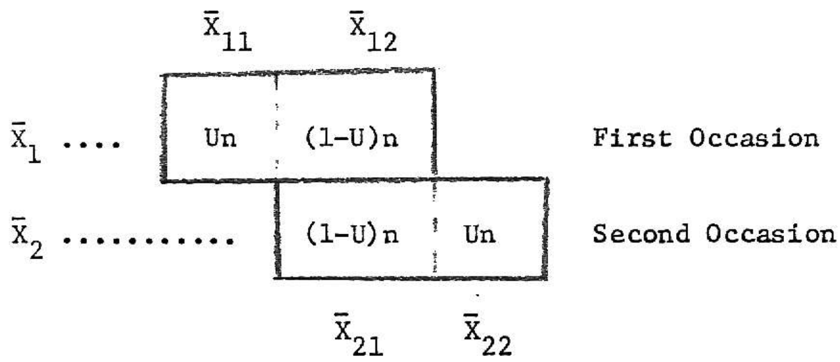
By using mean value from each occasion or averaged means from overall occasion, estimated total can be obtained as follows:

estimated total $\hat{T} = N\bar{X}$

ESTIMATION PROCEDURE IN TWO OCCASION SAMPLING

(a) Estimated Mean on the Second Occasion

From example 2, we generalize the case as follows:



Let $\bar{X}_{11}$ = sample mean from Un units on the first occasion

$\bar{X}_{12}$ = " " (1-U)n " " "

$\bar{X}_{21}$ = " " (1-U)n " on the second "

$\bar{X}_{22}$ = " " Un " " "

$\bar{X}_1$ = sample mean on the first occasion or mean per first occasion sampling unit which estimates μ_1, the population mean on the first occasion

$\bar{X}_2$ = sample mean on the second occasion which estimates μ_2, the population mean on the second occasion.

In this investigation, we are interested in the benefit of using a partial replacement sampling scheme on successive occasions. The estimate from the first occasion will be used as supplementary information. In other words, we wish to estimate by linear estimate.

D. Singh (4) presented a form of estimate as follows:

$$\bar{X}_2 = a\bar{X}_{11} + b\bar{X}_{12} + c\bar{X}_{21} + d\bar{X}_{22} \tag{1}$$

Where a, b, c, and d are arbitrary constants.

Now, for proper choice of these constants, $\bar{X}_2$ is an unbiased estimate of μ_2 since

$$E(\bar{X}_2) = aE(\bar{X}_{11}) + bE(\bar{X}_{12}) + cE(\bar{X}_{21}) + dE(\bar{X}_{22}) \qquad (2)$$

Consider $E(\bar{X}_{11})$, $E(\bar{X}_{12})$, $E(\bar{X}_{21})$ and $E(\bar{X}_{22})$:

$$\bar{X}_{11} = \frac{1}{Un} \sum_{i=1}^{Un} (X_{11})_i$$

$$E(\bar{X}_{11}) = \frac{1}{Un} \sum_{i=1}^{Un} E(X_{11})_i$$

$$= \frac{1}{Un}\, Un \sum_{i=1}^{N} \frac{1}{N} (X_{11})_i$$

$$= \frac{1}{N} \sum_{i=1}^{N} (X_{11})_i$$

$$= \mu_1$$

and

$$\bar{X}_{12} = \frac{1}{(1-U)n} \sum_{i=1}^{(1-U)n} (X_{12})_i$$

$$E(\bar{X}_{12}) = \frac{1}{(1-U)n} \sum_{i=1}^{(1-U)n} E(X_{12})_i$$

$$= \frac{1}{(1-U)n} (1-U)n \sum_{i=1}^{N} \frac{1}{N} (X_{12})_i$$

$$= \frac{1}{N} \sum_{i=1}^{N} (X_{12})_i$$

$$= \mu_1$$

We see that $E(\bar{X}_{11}) = E(\bar{X}_{12}) = \mu_1$ (3)

In the same way, $E(\bar{X}_{21}) = E(\bar{X}_{22}) = \mu_2$ (4)

Substitute (3) and (4) into (2):

$$E(\bar{X}_2) = (a+b)\mu_1 + (c+d)\mu_2$$

We see that $E(\bar{X}_2) = \mu_2$ iff $a+b = 0$

and $c+d = 1$.

This implies $(a+b) = 0$ and $(c+d) = 1$ make $\bar{X}_2$ an unbiased estimate of the population mean μ_2.

Substitute $a = -b$ and $c = 1-d$ into equation (2).

Then,

$$E(\bar{X}_2) = aE(\bar{X}_{11}) - aE(\bar{X}_{12}) + cE(\bar{X}_{21}) + (1-c)E(\bar{X}_{22})$$

$$= aE(\bar{X}_{11}-\bar{X}_{12}) + cE(\bar{X}_{21}) + (1-c)E(\bar{X}_{22}) \quad (5)$$

Therefore, the linear function which is an unbiased estimate of μ_2 is

$$\bar{X}_2 = a(\bar{X}_{11}-\bar{X}_{12}) + c(\bar{X}_{21}) + (1-c)\bar{X}_{22} \quad (6)$$

Apart from unbiased estimate, $\bar{X}_2$ will be the best linear unbiased estimate (BLUE) or not depend on the estimated variance on the second occasion.

(b) Estimated Variance on the Second Occasion

From (6), $V(\bar{X}_2) = V[a(\bar{X}_{11}-\bar{X}_{12}) + c(\bar{X}_{21}) + (1-c)\bar{X}_{22}]$

$$= a^2V(\bar{X}_{11}-\bar{X}_{12}) + c^2V(\bar{X}_{21}) + (1-c)^2V(\bar{X}_{22})$$

$$+ 2ac\ \mathrm{cov}[(\bar{X}_{11}-\bar{X}_{12}), \bar{X}_{21}]$$

$$+ 2a(1-c)\ \text{cov}[(\bar{X}_{11}-\bar{X}_{12}),\ \bar{X}_{22}]$$

$$+ 2c(1-c)\ \text{cov}[(\bar{X}_{21},\ \bar{X}_{22})] \qquad (7)$$

Consider each component in (7),

$$V(\bar{X}_{11}-\bar{X}_{12}) = V(\bar{X}_{11}) + V(\bar{X}_{12}) \qquad (8)$$

Proof.

$$V(\bar{X}_{11}-\bar{X}_{12}) = V(\bar{X}_{11}) + V(\bar{X}_{12}) + 2\ \text{cov}(\bar{X}_{11},\ \bar{X}_{12})$$

$$= V\left(\frac{1}{Un}\sum_{i=1}^{Un}(X_{11})_i\right) + V\ \frac{1}{(1-U)n}\sum_{i=1}^{(1-U)n}(X_{12})_i\Big)$$

$$- \frac{2}{Un(1-U)n}\ \text{cov}\left(\sum_{i=1}^{Un}(X_{11})_i,\ \sum_{i=Un+1}^{n}(X_{12})_i\right)$$

$$= \frac{1}{(Un)^2}\ V\left(\sum_{i=1}^{Un}(X_{11})_i\right) + \frac{1}{(1-U)^2n^2}\ V\left(\sum_{i=1}^{(1-U)n}(X_{12})_i\right)$$

$$- \frac{2}{Un(1-U)n}\ \text{cov}\left(\sum_{i=1}^{Un}(X_{11})_i,\ \sum_{i=Un+1}^{n}(X_{12})_i\right)$$

Ignoring finite population correction, namely $\frac{N-n}{N}$

$$V(\bar{X}_{11}-\bar{X}_{12}) = \frac{1}{(Un)^2}\ Un\ V(X_{11})_i + \frac{1}{(1-U)^2n^2}\ (1-U)n\ V(X_{12})_i$$

$$- 0$$

$$= \frac{1}{Un}\ S_1^2 + \frac{1}{(1-U)n}\ S_1^2$$

$$= V(\bar{X}_{11}) + V(\bar{X}_{12})$$

Q.E.D.

$$\text{cov}[(\bar{X}_{11} - \bar{X}_{12}), \bar{X}_{21}] = -\frac{\sigma_{12}}{(1-U)n} \tag{9}$$

<u>Proof</u>.

$$\text{cov}[(\bar{X}_{11} - \bar{X}_{12}), \bar{X}_{21}] = \text{cov}(\bar{X}_{11}, \bar{X}_{21}) - \text{cov}(\bar{X}_{12}, \bar{X}_{22})$$

$$= \frac{1}{Un(1-U)n} \text{cov}\left(\sum_{i=1}^{Un} (X_{11})_i, \sum_{i=Un+1}^{n} (X_{21})_i \right)$$

$$- \frac{1}{(1-U)^2 n^2} \text{cov}\left(\sum_{i=1}^{(1-U)n} (X_{12})_i, \sum_{i=1}^{(1-U)n} (X_{21})_i \right)$$

$$= 0 - \frac{1}{(1-U)^2 n^2} (1-U)n \,\text{cov}[(X_{12})_i, (X_{21})_i]$$

$$= -\frac{1}{(1-U)n} \sigma_{12}$$

<u>Q.E.D.</u>

Since $\bar{X}_{11}$, $\bar{X}_{12}$ and $\bar{X}_{22}$ or $\bar{X}_{11}$, $\bar{X}_{21}$ and $\bar{X}_{22}$ are independent with each other. Then,

$$\text{cov}[(\bar{X}_{11} - \bar{X}_{12}), \bar{X}_{22}] = 0 \tag{10}$$

and

$$\text{cov}[\bar{X}_{21}, \bar{X}_{22}] = 0 \tag{11}$$

Substitute (8), (9), (10) and (11) into (7), we get:

$$V(\bar{X}_2) = a^2[V(\bar{X}_{11}) + V(\bar{X}_{12})] + c^2\ V(\bar{X}_{21})$$

$$+ (1 - c)^2\ V(\bar{X}_{22}) - \frac{2ac\sigma_{12}}{(1-U)n}$$

$$= a^2\left(\frac{S_1^2}{Un} + \frac{S_1^2}{(1-U)n}\right) + c^2\left(\frac{S^2}{(1-U)n}\right)$$

$$+ (1 - c)^2\ \frac{S_2^2}{Un} - 2ac\ \frac{\sigma_{12}}{(1-U)n}$$

Where $S_1^2 = V(\bar{X}_{11}) = V(\bar{X}_{12})$

and $S_2^2 = V(\bar{X}_{21}) = V(\bar{X}_{22})$.

Let $\rho = \dfrac{\text{cov}(\bar{X}_1,\ \bar{X}_2)}{S_1 S_2}$

By assumption, $S_1^2 = S_2^2 = S^2$, then,

$$\text{cov}(\bar{X}_1,\ \bar{X}_2) = \rho S^2 = \sigma_{12}$$

Therefore,

$$V(\bar{X}_2) = \frac{a^2 S^2}{Un(1-U)} + \frac{c^2 S^2}{(1-U)n} + \frac{(1-c)^2 S^2}{Un} - \frac{2ac\rho S^2}{(1-U)n} \tag{12}$$

To find the relationship of a and c which minimize $V(\bar{X}_2)$, take the partial derivative of $V(\bar{X}_2)$ with respect to a and c, set them equal to zero. Thus,

$$\frac{\partial V(\bar{X}_2)}{\partial a} = \frac{2aS^2}{Un(1-U)} - \frac{2c\rho S^2}{(1-U)n} = 0$$

Multiply through by $\dfrac{Un(1-U)}{2S^2}$. The result will be

$$a - e\rho U = 0$$

Then $$c = \frac{a}{\rho U} \tag{13}$$

And $$\frac{\partial V(\bar{X}_2)}{\partial c} = \frac{2cS^2}{(1-U)n} - \frac{2(1-c)S^2}{Un} - \frac{2a\rho S^2}{(1-U)n} = 0$$

Multiply through by $\dfrac{Un(1-U)}{2S^2}$, one obtains,

$$cU - (1 - c)(1 - U) - a\rho U = 0$$

Substitute the value of c in (13), and obtain

$$\frac{aU}{U\rho} - \left(1 - \frac{a}{U\rho}\right)\left(1 - U\right) - aU\rho = 0$$

$$\frac{aU}{U\rho} - \left(1 - U - \frac{a}{U\rho} + \frac{aU}{U\rho}\right) - aU\rho = 0$$

$$aU - U\rho + U^2\rho + a - aU - aU^2\rho^2 = 0$$

$$a - aU^2\rho^2 = U\rho - U^2\rho$$

$$a = \frac{U\rho(1-U)}{1 - U^2\rho^2} \tag{14}$$

Substitute the value of a in (14) back to (13), and obtain

$$c = \frac{U\rho(1-U)}{U\rho(1 - U^2\rho^2)}$$

$$c = \frac{1-U}{1 - U^2\rho^2} \tag{15}$$

The values of a in (14) and c in (15) make

$$\bar{X}_2 = a(\bar{X}_{11} - \bar{X}_{12}) + c(\bar{X}_{21}) + (1 - c)\bar{X}_{22}$$

The BLUE value of μ_2.

Therefore, the BLUE value of μ_2 becomes:

$$\bar{X}_2 = \frac{(1-U)U\rho}{1-U^2\rho^2}(\bar{X}_{11} - \bar{X}_{12}) + \frac{(1-U)\bar{X}_{21}}{1-U^2\rho^2} + \left(1 - \frac{1-U}{1-U^2\rho^2}\right)\bar{X}_{22} \tag{16}$$

For estimated variance, substitute the values of a and c in terms of U and ρ into (12), we obtain

$$V(\bar{X}_2) = \frac{(1-U)U\rho^2S^2}{(1-U^2\rho^2)n} + \frac{(1-U)S^2}{(1-U^2\rho^2)n}$$

$$+ \left(\frac{1-U^2\rho^2-1+U}{1-U^2\rho^2}\right)^2 \frac{S^2}{Un} - \frac{2(1-U)U\rho^2S^2}{n(1-U^2\rho^2)^2} \tag{17}$$

$$V(\bar{X}) = \frac{(1-U)\rho^2S^2U}{n(1-U^2\rho^2)^2} + \frac{(1-U)S^2}{n(1-U^2\rho^2)^2}$$

$$+ \frac{U(1-U\rho^2)^2S^2}{n(1-U^2\rho^2)^2} - \frac{2(1-U)U\rho^2S^2}{n(1-U^2\rho^2)^2}$$

$$= \frac{S^2}{n(1-U^2\rho^2)}\left(\frac{\rho^2(1-U)U+1-U+U(1-2U\rho^2+U^2\rho^2)-2U\rho^2+2U^2\rho^2}{1-U^2\rho^2}\right)$$

$$= \frac{S^2}{n(1-U^2\rho^2)}\left(\frac{1-U^2\rho^2-U\rho^2+U^3\rho^4}{1-U^2\rho^2}\right)$$

$$= \frac{S^2}{n(1-U^2\rho^2)} \left\{\frac{1-U^2\rho^2 - U\rho^2(1-U^2\rho^2)}{1-U^2\rho^2}\right\}$$

$$= \frac{S^2}{n(1-U^2\rho^2)} \left\{\frac{(1-U^2\rho^2)(1-U\rho^2)}{1-U^2\rho^2}\right\}$$

$$= \frac{S^2}{n} \frac{(1-U\rho^2)}{(1-U^2\rho^2)} \tag{18}$$

COMPARISON OF PARTIAL REPLACEMENT WITH FIXED SAMPLE AND INDEPENDENT SAMPLES.

Consider the value of the minimum variance as given in (18):

$$V(\bar{X}_2) = \frac{S^2}{n} \left\{\frac{1-U\rho^2}{1-U^2\rho^2}\right\}$$

Where U is the replacement rate for the second occasion, $0 < U < 1$.

In fixed sample scheme, $U = 0$,

$$V(\bar{X}_2) = \frac{S^2}{n} \left\{\frac{1-0\rho^2}{1-0\rho^2}\right\}$$

$$= \frac{S^2}{n}, \tag{19}$$

and in independent sample, $U = 1$,

$$V(\bar{X}_2) = \frac{S^2}{n} \left\{\frac{1-\rho^2}{1-\rho^2}\right\}$$

$$= \frac{S^2}{n} \tag{20}$$

Therefore, the difference between partial replacement and the two schemes is $\frac{1-U\rho^2}{1-U^2\rho^2}$,

where $$0 < \frac{1-U\rho^2}{1-U^2\rho^2} < 1, \qquad \rho \neq 0.$$

This implies $V(\bar{X}_2)$ of partial replacement scheme is less than $\frac{S^2}{n}$. In other words, partial replacement in two occasions sampling gives a variance that is less than fixed and independent samples.

<u>Example</u> For $U = \frac{1}{2}$ and $\rho = 0.9$,

$$V(X_2) = \frac{S^2}{n}\left(\frac{1-\frac{1}{2}(0.81)}{1-\frac{1}{4}(0.81)}\right)$$

$$= \frac{2.38}{3.19}\frac{S^2}{n}$$

$$= 0.74\frac{S^2}{n} \quad \text{which is less than } \frac{S^2}{n}.$$

Therefore, a partial replacement scheme gives a more efficient estimate than either a fixed sample or independent samples scheme.

OPTIMUM VALUE OF U IN THE TWO OCCASIONS SAMPLING.

Since $0 < U < 1$, we can assign any value of U within 0 and 1 for our replacement rate on the second occasion. The optimum value of U yields the most efficient estimate.

Theoretically, we can find an optimum value of U when the correlation coefficient of the characteristic under study is known. We minimize $V(\bar{X}_2)$ by taking partial derivative of $V(\bar{X}_2)$ with respect to U,

$$\frac{\partial V(\bar{X}_2)}{\partial U} = \frac{S^2}{n}\left(\frac{(1-U^2\rho^2)(-\rho^2)-(1-U\rho^2)(-2U\rho^2)}{(1-U^2\rho^2)^2}\right.$$

$$= -\rho^2 + U^2\rho^4 + 2U\rho^2 - 2U^2\rho^4 = 0$$

$$- U^2\rho^2 + 2U - 1 = 0$$

$$U = \frac{-2 \pm \sqrt{4-4(-\rho^2)(-1)}}{-2\rho^2}$$

$$U_{opt.} = \frac{1 - \sqrt{1-\rho^2}}{\rho^2} \tag{21}$$

Ignore $\frac{1 + \sqrt{1-\rho^2}}{\rho^2}$, since it is greater than 1.

ESTIMATION OF CORRELATION COEFFICIENT

Before one assign the value of a replacement rate U or calculates the optimum value of U as in (21), it is necessary to know the value of the correlation coefficient, ρ.

In practice, the value of a population correlation coefficient of the same sample on two-occasion sampling is rarely known. We estimate ρ by r, the sample correlation coefficient.

By definition, $r = \frac{\text{estimated cov }(X_1, X_2)}{\text{estimated var}(X_1)\text{var}(X_2)}$

$$= \frac{\sum_{i=1}^{n'} (X_{1i}-\bar{X}_1)(X_{2i}-\bar{X}_2)}{\sum_{i=1}^{n'} (X_{1i}-\bar{X}_1) \sum_{i=1}^{n'} (X_{2i}-\bar{X}_2)}$$

X_{1i} = the i-th observation on the first occasion,

i = 1,2,...,n'

X_{2i} = the i-th observation on the second occasion,

i = 1,2,...,n'

n' = number of repeated observation on the second occasion.

From (21), substitute ρ by r, one obtains estimated replacement rate.

ACKNOWLEDGMENT

I am very much indepted to Professor A. M. Feyerherm for his advice and assistance during the preparation of this report.

I wish to express my sincere gratitude to Professor H. C. Fryer for his kind encouragement given to me during my graduate work. I shall never forget my arrival in Manhattan when I was greatly honored by Dr. and Mrs. Fryer personally greeting me at the airport.

I also wish to acknowledge the financial support provided by AID program during my two years of graduate study in the United States.

BIBLIOGRAPHY

(1) Jessen, R.J., "Statistical Investigation of a Sample Survey for Obtaining Farm Facts," (Iowa Agricultural Experiment Station Research Bulletin, No. 304, 1950)

(2) Patterson, H.D., Sampling on Successive Occasions with Partial Replacement of Units. (Journal of Royal Statistical Series B. 12, 1960) p. 241-255

(3) Rao, J.N.K., Rotation Design for Sampling on Repeated Occasions. (American Statistical Association Journal, June 1964) p. 492-509

(4) Singh, D., Estimates in Successive Sampling Using a Multistage Design. (American Statistical Association Journal, March 1968) p. 99-112

(5) Yates, Frank F.R.S., Sampling Methods for Censuses and Surveys. (London: Charles Griffin & Co, Ltd., 1965)

SAMPLING ON TWO OCCASIONS
WITH PARTIAL REPLACEMENT OF UNITS

by

NANTIT SUPAMONGKON

B.A., Chulalongkorn University, 1964

AN ABSTRACT OF A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1970

ABSTRACT

This paper is a report of a repeated sampling procedure for two occasions using unistage sampling designs. Three sampling schemes namely; fixed sample, independent samples and partial replacement were presented. It was found that a partial replacement scheme was superior to the other two schemes.

By using a partial replacement scheme, the estimation of mean and variance on the last occasion using the estimate from the first occasion as a supplementary information was presented with a comparison of the efficiencies gained from these three schemes. It was concluded that partial replacement gave a more efficient estimate than either a fixed sample or independent sample scheme.

The estimation procedure was developed under the stationary assumption; that is, population size was fixed during the process of estimation and the variances were assumed independent of time.