A SIMPLE DESIGN IN SAMPLING WITH VARYING PROBABILITIES

By A. R. Sen

Indian Council of Agricultural Research, New Delhi

Introduction

The theory of selecting a combination of n primary sampling units (p.s.u.s.) from a stratum with probability proportional to sum of the measures of the units in the combination was developed independently by Midzuno[2] and the present author[3] who derived a simple expression for the unbiased estimate of the population total. The present author[3] further derived an expression for the unbiased estimate of the variance of the estimated total which was later shown[4] to have generally no practical application as it can assume negative values. Recently Horvitz and Thomson[1] developed a scheme for obtaining unbiased estimates of the population total and of the variance of the estimate when n p.s.u.s. are selected without replacement from a stratum using arbitrary probabilities of selection for elements remaining prior to each draw. The method is not practicable since the expression for the estimates of the total and of the variance of the estimated total become extremely unwieldy when it is required to select more than two p.s.u.s. Further, the estimate of the variance has no practical application in this case as it may assume negative values when the sample size exceeds two. One possible solution would be to carry stratification to the stage when two p.s.u.s. could be selected from each stratum. This is, however, not generally practicable because increasing the number of strata beyond a certain point may not be profitable or even feasible from administrative view-point.

2. The present design is mainly intended to meet the disadvantage in the Midzuno and Sen case when more than two p.s.u.s. are to be selected from a stratum and provides simple expressions for the estimates of the total and of the variance of the estimated total. The principle contribution in this paper is the development of the theory of sampling with probability proportional to a measure of size (p.p.s.) when $2n$ p.s.u.s. are selected from a stratum in two groups of n each, such that either the elements within a group are selected with p.p.s. and without replacement as in Horvitz and Thomson's scheme or the first unit is selected with p.p.s. and the rest with equal probability as in Midzuno and Sen's method and the groups are selected

with replacement. The theory, thus, developed has been generalised for the case when mn p.s.u.s. are to be selected in m groups with n elements in each group. Efficient sampling systems have been derived for two-stage sampling when the groups may have overlapping p.s.u.s. The theory has been applied to the case when 4 p.s.u.s. are to be selected in groups of two each with replacement from a stratum (*a*) when the units in a group are selected with p.p.s. and without replacement, (*b*) when one of the units in a group is selected with p.p.s. and the other with equal probability from the remaining units. The efficiencies of the sampling systems for the selection procedures (i) when the estimates are made according to Horvitz and Thomson, (ii) when the estimate is based on Midzuno and Sen scheme are compared with those of systems where (iii) one of the units is selected with p.p.s. and the remaining three units with equal probability but without replacement from the stratum, (iv) the units are selected with p.p.s. and with replacement from the stratum, (v) the units are selected at random with equal probability but without replacement from the stratum and (vi) the stratum is divided into two strata and two units are selected from each new stratum as in (i).

SAMPLING SYSTEMS

3. Consider for simplicity one stratum only. The results obtained for one stratum can be generalised for K strata. Let $Y_1, \cdots, Y_N$ be a population of N p.s.u.s. comprising the stratum with respective measures of size proportional to $X_1, \cdots, X_N$. In particular, the X_i, s may be previous census values of the characteristic known exactly. Let M_i be the number of secondary or sub-sampling units (s.s.u.s.) in the ith p.s.u. Also let Y_{ij} be the value of some characteristic Y of the jth s.s.u. in the ith p.s.u. The population total

$$Y = \sum_{i=1}^{N} \sum_{j=1}^{M_i} Y_{ij} \tag{1}$$

is to be estimated from a sample of $2n$ p.s.u.s. The basic sampling design consists of two stages of sampling (*a*) the $2n$ p.s.u.s. are selected from the stratum in two groups so that the groups are selected with replacement but the p.s.u.s. within a group are selected with p.p.s. and without replacement. For simplicity, we shall assume that the groups are equal in size, *i.e.*, each group consisting of n p.s.u.s.; and from each of the $2n$ p.s.u.s., s.s.u.s. are selected at random and without replacement. It may be that one or more of the p.s.u.s. are common to the two groups in which case some of the s.s.u.s. may be repeated in a p.s.u. resulting in loss of efficiency. In such a case another sample design will be considered where twice the number of s.s.u.s. are selected from a p.s.u. if the p.s.u. happens to be selected again. This device is not new and has been employed by Sukhatme and Narain[6] in another situation. Two methods of estimation will be employed for each of the two sample designs resulting in four sampling systems A, B, C and D which are as follows:—

A. (i) *p.s.u. selection.*—The first n p.s.u.s. are selected without replacement using arbitrary probabilities of selection for each draw. These are then replaced and another set of n p.s.u.s. are drawn similarly thus selecting in all $2n$ p.s.u.s.

(ii) *s.s.u. selection.*—A number of sub-sampling units (m_i) are selected independently and at random from the ith p.s.u., if it is in the sample.

(iii) *Estimate of the total.*—

$$\hat{Y}_1 = \tfrac{1}{2}\left[\sum_{i=1}^{n} \frac{y_i'}{P_{(u_i)}} + \sum_{j=1}^{n} \frac{y_j'}{P_{(u_j)}}\right] \tag{2}$$

where $P(u_i)$ and $P(u_j)$ denote the *apriori* probabilities that the ith and the jth p.s.u.s. are included in the two groups of n p.s.u.s. each. Also let y_i' and y_j' be unbiased estimates of the totals Y_i and Y_j respectively. In particular, when the first element is drawn with arbitrary probability and the subsequent p.s.u.s. with equal probability and without replacement the estimate is given by (2) where

$$P_{(u_i)} = \frac{N-n}{N-1}\, p_i + \frac{n-1}{N-1} \tag{3}$$

where

$$\sum_{i=1}^{N} p_i = 1$$

B. (i) *p.s.u. selection.*—Same as in A (i).

(ii) *s.s.u. Selection.*—A number of s.s.u.s. (m_i) are selected at random from the ith p.s.u., if it is in the sample. In case a p.s.u. is reselected, twice the number of s.s.u.s. are to be selected from the p.s.u. at random and without replacement.

(iii) *Estimate of the total.*—

$$\hat{Y}_2 = \tfrac{1}{2}\left[\sum_{i=1}^{n} \frac{y_i'}{P_{(u_i)}} + \sum_{j=1}^{n} \frac{y_j'}{P_{(u_j)}}\right] \tag{4}$$

where y_i' and y_j' are unbiased estimates of Y_i and Y_j.

C. (i) *p.s.u. selection.*—The first p.s.u. is selected with p.p.s. and the remaining $n-1$ units are selected with equal probability but without replacement. The n units thus selected are then replaced and another set of n units are selected in exactly the same way from the stratum thus selecting in all $2n$ p.s.u.s.

(ii) *s.s.u. selection.*—Same as in A (ii).

(iii) *Estimate of total.*—

$$\hat{Y}_3 = \tfrac{1}{2}\left[\frac{\Sigma_1 y_i'}{\Sigma_1 X_i} X + \frac{\Sigma_2 y_j'}{\Sigma_2 X_j} X\right] \tag{5}$$

where Σ_1 and Σ_2 are summations over the two groups of n p.s.u.s. each, selected with replacement in accordance with (i) above, X_i is the known measure of size of the ith unit and $X = \sum_{i=1}^{N} X_i$.

D. (i) *p.s.u. selection.*—Same as in C (i).

(ii) *s.s.u. selection.*—Same as in B (ii).

(iii) *Estimate of the total.*—

$$\hat{Y}_4 = \tfrac{1}{2}\left[\frac{\Sigma_1 y_i'}{\Sigma_1 X_i} X + \frac{\Sigma_2 y_j'}{\Sigma_2 X_j} X\right] \tag{6}$$

where y_i' and y_j' are unbiased estimates of Y_i and Y_j.

It is easy to see that $\hat{Y}_1$ and $\hat{Y}_2$ are unbiased estimates of the population total Y since each of the expressions $\sum_{i=1}^{n} \frac{y_i'}{P_{(u_i)}}$ and $\sum_{j=1}^{n} \frac{y_j'}{P_{(u_j)}}$ is an unbiased estimate of Y. Similarly $\hat{Y}_3$ and $\hat{Y}_4$ are unbiased estimates of the population total Y.

4. We will now derive expressions for the variances of the estimates $\hat{Y}_1$, $\hat{Y}_2$, $\hat{Y}_3$ and $\hat{Y}_4$ and expressions for estimates of the variances. Let r_1 be the number of p.s.u.s. common to each of the two groups for the sampling systems A and B and r_2 the number common for the sampling systems C and D.

$$V(\hat{Y}_1) = E(\hat{Y}_1 - Y)^2$$

$$= \frac{V_1}{2} + \tfrac{1}{2} \sum_{i}^{N} \frac{M_i (M_i - m_i)}{P_{(u_i)}} \frac{\sigma_i^2}{m_i} \tag{7}$$

where

$$V_1 = \sum_{i=1}^{N} \frac{Y_i^2}{P_{(u_i)}} + \sum\sum_{i \neq j}^{N} \frac{P_{(u_i u_j)}}{P_{(u_i)} \cdot P_{(u_j)}} Y_i Y_j - Y^2$$

Again

$$V(\hat{Y}_2) = E(\hat{Y}_2 - Y)^2$$

$$= \frac{V_1}{2} + \frac{E}{4} \left[\frac{4r_1}{n} \sum_{i}^{N} \frac{M_i (M_i - 2m_i)}{P_{(u_i)}} \frac{\sigma_i^2}{2m_i} + 2 \left(\frac{n - r_1}{n} \right) \sum_{i}^{N} \frac{M_i (M_i - m_i)}{P_{(u_i)}} \frac{\sigma_i^2}{m_i} \right]$$

$$= \frac{V_1}{2} + \tfrac{1}{2} \sum_{i=1}^{N} \frac{M_i (M_i - m_i)}{P_{(u_i)}} \frac{\sigma_i^2}{m_i} - \frac{\bar{r}_1}{2n} \sum_{i}^{N} \frac{M_i \sigma_i^2}{P_{(u_i)}} \tag{8}$$

where

$$E(r_1) = \bar{r}_1$$

$$V(\hat{Y}_3) = \frac{V_2}{2} + \frac{E}{2} \left[\sum_{i}^{n} M_i (M_i - m_i) \frac{\sigma_i^2}{m_i} \frac{X^2}{(\Sigma X_i)^2} \right] \tag{9}$$

where

$$V_2 = \frac{1}{\binom{N-1}{n-1}} \sum\sum_{i<j} \cdots \sum_{<l} \frac{\left(\sum_{i}^{n} y_i \right)^2}{\sum_{i}^{n} X_i} X - Y^2$$

and $\sum\sum_{i<j} \cdots \sum_{<l}$ stands for summation over all combinations $\binom{N}{n}$ of terms of n units $(i, j, \cdots, l)$ and $X = \sum_{i=1}^{N} X_i$

$$V(\hat{Y}_4) = E(\hat{Y}_4 - Y)^2$$

$$= \frac{V_2}{2} + E \left[\sum_{i=1}^{r_2} M_i (M_i - 2m_i) \frac{\sigma_i^2}{2m_i} \right.$$

$$+\frac{+\sum_{j=1}^{n-r_2} M_j(M_j-m_j)}{\left(\sum^{n} X_i\right)^2}\frac{\sigma_j^2}{m_j}\cdot\frac{X^2}{2}$$

$$\left.+\tfrac{1}{2}\sum_{k=1}^{r_2} M_k(M_k-2m_k)\frac{\sigma_k^2}{2m_k}\cdot\frac{X}{\Sigma X_i}\cdot\frac{X}{\Sigma X_j}\right]$$

where K runs over r_2 p.s.u.s. which are common to the two groups selected with replacement

$$=\frac{V_2}{2}+\frac{E}{2}\left[\frac{\sum_{i=1}^{n} M_i(M_i-m_i)\frac{\sigma_i^2}{m_i}-\sum_{k=1}^{r_2}\frac{M_k^2\sigma_k^2}{2m_k}}{\left(\sum_i^n X_i\right)^2}X^2\right.$$

$$\left.+\sum_{k=1}^{r_2} M_k(M_k-2m_k)\frac{\sigma_k^2}{2m_k}\cdot\frac{X^2}{\sum_i^n X_i\sum_j^n X_j}\right]$$

$$=V(\hat{Y}_3)-\frac{E}{2}\left[\sum_{k=1}^{r_2}\frac{M_k^2\sigma_k^2}{4m_k}\left(\frac{X}{\sum_i^n X_i}-\frac{X}{\sum_j^n X_j}\right)^2\right.$$

$$\left.+\sum_{k=1}^{r_2}\frac{M_k\sigma_k^2}{\left(\sum_i^n X_i\right)\left(\sum_j^n X_j\right)}X^2\right] \qquad (10)$$

COMPARISON OF EFFICIENCY

5. From (7) and (8) it is clear that

$$V(\hat{Y}_2) < V(\hat{Y}_1)$$

so that system B is more efficient than system A.

Also

$$V(\hat{Y}_1)-V(\hat{Y}_2)=\frac{\bar{r}_1}{2n}\sum_i^N\frac{M_i\sigma_i^2}{P_{(u_i)}} \qquad (11)$$

which is independent of the amount of sub-sampling in the p.s.u.s. selected.

In particular if $P_{(u_i)}=n/N$, *i.e.*, when the p.s.u.s. are selected at random and without replacement (11) reduces to

$$\frac{\bar{r}N}{2n^2} \sum_i^N M_i \sigma_i^2 \tag{12}$$

where $\bar{r}$ is the expected value of the p.s.u.s. common to two groups when the p.s.u.s. within a group are selected at random with equal probability and without replacement and the groups are selected with replacement.

We have from (10)

$$V(\hat{Y}_3) - V(\hat{Y}_4) = \frac{E}{2}\left[\sum_{k=1}^{r_2} \frac{M_k{}^2\sigma_k{}^2}{4m_k}\left(\frac{X}{\sum_i^n X_i} - \frac{X}{\sum_j^n X_j}\right)^2 + \sum_{k=1}^{r_2} \frac{M_k\sigma_k{}^2}{\left(\sum_i^n X_i\right)\left(\sum_j^n X_j\right)} X^2\right] \tag{13}$$

Hence $V(\hat{Y}_3) > V(\hat{Y}_4)$ so that system D is more efficient than system C.

In the special case where $X_1 = X_2 = \cdots = X/N$, *i.e.*, when the p.s.u.s. are selected at random and without replacement (13) reduces to

$$\frac{\bar{r}N}{2n^2} \sum_i^N M_i \sigma_i^2$$

which is the same as (12).

Values of $E(r_1)$ and $E(r_2)$

6. The values of $E(r_1)$ and $E(r_2)$ will be given for the important practical case when four p.s.u.s. are selected in groups of two with replacement. It is easy to see that the probabilities for 0, 1, 2 overlapping units are:

$$\sum_{i<j}\sum [1 - P_{(u_i)} - P_{(u_j)} + P_{(u_iu_j)}]\, P_{(u_iu_j)},$$

$$\sum_{i<j}\sum [P_{(u_i)} + P_{(u_j)} - 2\, P_{(u_iu_j)}]\, P_{(u_iu_j)}$$

and

$$\sum_{i<j}\sum P^2_{(u_iu_j)}$$

Hence

$$E(r_1) = \sum_{i<j}\sum P_{(u_iu_j)}\, [P_{(u_i)} + P_{(u_j)}] \tag{14}$$

When the first p.s.u. is selected with p.p.s. and the second with equal probability from among the remaining units

$$P_{(ui)} = \frac{(N-2)\, p_i + 1}{N-1}$$

$$P_{(u_i u_j)} = \frac{(p_i + p_j)}{N-1} \qquad (15)$$

where

$$\sum_{i}^{N} p_i = 1$$

Substituting the values of $P_{(u_i)}$ and $P_{(u_i u_j)}$ from (15) in (14) we have

$$E(r_2) = \sum_{i<j}\sum \frac{(p_i + p_j)}{(N-1)^2} \{(p_i + p_j)(N-2) + 2\} \qquad (16)$$

Estimates of Error

7. An unbiased estimate of (7) is given by

$$\tfrac{1}{4}\left[\sum_{i=1}^{n} \frac{y_i'}{P_{(u_i)}} - \sum_{j=1}^{n} \frac{y_j'}{P_{(u_j)}}\right]^2 \qquad (17)$$

when the s.s.u.s. are selected according to A (ii), *i.e.*, independently and at random from each p.s.u. irrespective of the fact a p.s.u. is repeated or not.

When the s.s.u.s. are selected according to B (ii), (17) is a biased estimate of (8). An unbiased estimate of (8) is given by

$$\tfrac{1}{4}\left[\sum_{i=1}^{n-r} \frac{y_i'}{P_{(u_i)}} - \sum_{j=1}^{n-r} \frac{y_j'}{P_{(u_j)}}\right]^2 + \sum_{k=1}^{r_1} \frac{M_k (M_k - 2m_k)\, s_k^2}{2m_k . P^2_{(u_k)}} \qquad (18)$$

An unbiased estimate of (9) is given by

$$\tfrac{1}{4}\left[\frac{\sum_1 y_i'}{\sum_1 X_i} X - \frac{\sum_2 y_j'}{\sum_2 X_j} X\right]^2 \qquad (19)$$

where the s.s.u.s. are selected according to B (ii), *i.e.*, independently and at random from each p.s.u. in the sample, whether the p.s.u. is reselected or not.

When the s.s.u.s. are selected according to D (ii), (19) is a biased estimate of (10). An unbiased estimate of (10) is given by

$$\frac{1}{4}\left[\frac{\sum_1 y_i'}{\sum_1 X_i} X - \frac{\sum_2 y_j'}{\sum_2 X_j} X\right]^2$$

$$+ \sum_{k=1}^{r_2} \frac{M_k (M_k - 2m_k)}{2m_k} \cdot \frac{s_k^2}{\sum_1 X_i} \cdot \frac{X^2}{\sum_2 X_j} \qquad (20)$$

General Case

8. We will now deal with the more general case when mn p.s.u.s. are selected in m groups with n p.s.u.s. in each group, the groups being selected with replacement. We may for instance plan the fieldwork with different fieldworkers in different groups to avoid any possible bias due to correlation between the errors of measurements of the different fieldworkers.

Consider a multi-stage sampling system in which a group of n p.s.u.s. are selected as the first stage units with p.p.s. and without replacement and independent samples are taken at the second and subsequent stages for each drawing. Let the m groups be selected with replacement. Then if $\hat{a}_1, \cdots, \hat{a}_m$ be the unbiased estimates of the total we have since $\hat{a}_1, \cdots, \hat{a}_m$ are independent estimates of the total, an unbiased estimate of the total is given by

$$\hat{a} = \frac{\sum_i^m \hat{a}_i}{m} \qquad (21)$$

An unbiased estimate of error is easily seen to be

$$\frac{\sum_i^m (\hat{a}_i - \hat{a})^2}{m(m-1)} \qquad (22)$$

An unbiased estimate of error is also provided by pooling the independent estimates of error for each sample of n p.s.u.s. according to Yates and Grundy[7] or Sen[5] method. This will have practical application only for the case $n = 2$ since for $n > 2$ the estimate may assume negative values. For large m the estimate of error (22) will be computationally simpler than that obtained by averaging the independent estimates of error (5, 7). For small m, however, the latter method may be more efficient in view of larger degrees of freedom for the estimate of error. This point needs further investigation.

The case of m-stages with samples of 2 at each stage when the p.s.u.s. are selected as in Horvitz and Thomson's scheme is of some practical interest and follows as special cases of the results (21) and (22) derived above where $\hat{a}_i$ ($i = 1, \cdots, m$) are based on samples of 2 p.s.u.s.

Consider now the sampling system where the selection of the units at the second and subsequent stages from a p.s.u. is not independent of that from another. For example, for a two-stage sampling system let λ times the assigned number of s.s.u.s. be selected from a p.s.u. if the p.s.u. is repeated λ times ($\lambda \leqslant m$). An unbiased estimate of the total is still given by (21). Expression (22), however, provides only a biased estimate of the error variance.

Application

9. The population under study consists of acreage under irrigated wheat for the various patwari circles of pargana Mohanlalganj in District Lucknow, the data being given in Table I. Column 2 gives the acreage under irrigated wheat in 1950–51 known from a complete census. The correlation between the acreages for 1950–51 and 1951–52 for the population is 0·9.

TABLE I

Circle	Acreage under irrigated wheat	
	(1950–51)	(1951–52)
1	2	3
1	119	190
2	102	162
3	151	302
4	283	429
5	228	291
6	488	524
7	410	414
8	447	440
9	440	468
10	406	515
11	457	432
12	398	458
13	460	510
14	339	449
15	245	383
16	371	466
17	385	472
18	465	677
Total ..	6194	7582

The object is to make a preharvest estimate of the total acreage under wheat in 1951–52 of pargana Mohanlalganj by a complete enumeration of the acreage under the crop from a sample of villages. In practice we may have two-stage sampling systems with circles as first stage and villages as second stage units. The efficiencies of various unbiased sampling systems will be examined by comparing the exact between circle component of variance of the estimate for the case where 4 circles are sampled for each sampling system. For this, exact values of $(C.V.)^2\times 10^3$ where $(C.V.)^2$ is the true between circle component of variance divided by square of the total acreage for 1951–52, are presented in Table II.

TABLE II

$(C.V.)^2\times 10^3$ *for the between p.s.u. Component of Error*

Sl. No.	Sampling system		$(C.V.)^2\times 10^3$	Relative efficiency
	Method of selection	Estimate		
1	Simple random	$N\bar{y}$	20	42
2	First p.p.s. and rest equal without replacement	$\sum_{i=1}^{4}\frac{y_i}{P_{(ui)}}$ where $P_{(ui)}=\frac{N-4}{N-1}p_i+\frac{3}{N-1}$	11	78
3	p.p.s. with replacement	$\frac{1}{4}\sum_{i}^{4}\frac{y_i}{p_i}$	8	100
4	p.s.u.s. selected according to proposed scheme (C)	$\frac{1}{2}\left[\frac{\sum_i^2 y_i}{\sum_i X_i}X+\frac{\sum_j^2 y_j}{\sum_j X_j}X\right]$	7	114
5	p.s.u.s. selected according to proposed scheme (A)	$\frac{1}{2}\left[\sum_{i=1}^{2}\frac{y_i}{P_{(ui)}}+\sum_{j=1}^{2}\frac{y_j}{P_{(uj)}}\right]$	6	142
6	Stratum divided into 2 strata and 2 p.s.u.s. selected from each stratum according to Horvitz and Thomson's scheme	$\sum_i^2\frac{y_i}{P^1_{(ui)}}+\sum_j^2\frac{y_j}{P^2_{(uj)}}$ where $P^1_{(ui)}$ and $P^2_{(uj)}$ stand for strata 1 and 2 respectively	4	218

A comparison of the efficiencies of the sampling systems would show that simple random sampling is least efficient compared to other systems where the data of 1950–51 are utilized as ancillary information. Sampling system (6) is most efficient and is twice as efficient as system (3) when selection is made with p.p.s. and with replacement. In the sampling system (6) the circles are first divided into 2 homogeneous strata. The strata are determined by first ranking the circles on the basis of acreage under irrigated wheat during 1950–51, the circle with the maximum acreage receiving rank one. The first 9 circles form one of the strata and the remaining 9 form the other. Two p.s.u.s. are then selected from each stratum with p.p.s. according to Horvitz and Thomson's scheme. System (5) when the p.s.u.s. are selected in two groups with replacement according to Horvitz and Thomson's scheme is about $1\frac{1}{2}$ times as efficient as system (3). System (4) when the p.s.u.s. are selected in two groups with replacement according to Midzuno and Sen scheme is more efficient than system (3) and about $1\frac{1}{2}$ times as efficient as system (2) when the units are selected without replacement according to Midzuno and Sen scheme. It is interesting to note that system (3) is about $1\frac{1}{3}$ as efficient as system (2) which suggests that for high correlations between the auxiliary variate and the variate under study p.p.s. with replacement is superior to selecting a number of units without replacement, the first being selected with p.p.s. and the subsequent units with equal probability as in Midzuno and Sen scheme.

SUMMARY

Theory has been developed for the unbiased system when $2n$ p.s.u.s. are selected from a stratum in two groups of n each, the p.s.u.s. within a group being selected either in accordance with Midzuno and Sen or Horvitz and Thomson's scheme and the groups selected with replacement. Simple expressions for the unbiased estimates of the total and of the variance of the estimated total are derived in such a case. The theory has been generalised for the selection of mn p.s.u.s. from a stratum in m groups with replacement with n p.s.u.s. in each group.

2. Efficient sampling systems have been derived for two-stage sampling when the groups may have overlapping p.s.u.s. It is shown that the sampling system with twice the number of s.s.u.s. ($2m_i$) selected from a p.s.u. when the p.s.u. is repeated is more efficient than selecting m_i s.s.u.s. independently each time a p.s.u. is reselected.

3. The theory has been applied to the selection of 4 p.s.u.s. in groups of two each from a stratum, where one of the p.s.u.s. in a group is selected with p.p.s. and the other with equal probability or where **both** the units are selected with p.p.s. and without replacement, the

groups being selected with replacement and the efficiencies are compared with those of existing unbiased systems.

ACKNOWLEDGEMENT

The author wishes to express his appreciation to Dr. V. G. Panse for his useful comments which helped to improve the paper. The writer is also thankful to Shri Daroga Singh for certain suggestions in the final preparation of this paper.

REFERENCES

1. Horvitz, D. G. and Thomson, D. J. "A generalisation of sampling without replacement from a finite universe," *J.A.S.A.*, 1952, **47**, 663-85.
2. Midzuno, H. .. "On the sampling system with probability proportional to sums of sizes," *Annals of the Inst. of Stat. Math.* (*Tokyo*), 1952, **3**, 99-107.
3. Sen, A. R. .. "Further developments of the theory and application of the selection of primary sampling units with special reference to N.C. Agricultural Population," *Thesis*, Library, N.C. State College, 1952.
4. ———— .. "On the selection of n primary sampling units from a stratum structure ($n \geqslant 2$)," *Annals of Math. Stats.*, 1955, **26**, 744-51.
5. ———— .. "On the estimate of the variance in sampling with varying probabilities," *J. Ind. Soc. Agr. Stat.*, 1953, **5**, 119-27.
6. Sukhatme, P. V. and Narain, R. D. "Sampling with replacement," *Ibid.*, 1952, **4**, 42.
7. Yates, F. and Grundy, P. M. "Selection without replacement from within strata with probability proportional to size," *J. Roy. Stat. Soc. Ser., B*, 1953, **15**, 235-61.