ON THE THEORY OF SYSTEMATIC SAMPLING, II

By William G. Madow[1]

Institute of Statistics, University of North Carolina

1. Summary and introduction. In an earlier paper,[2] [1] an approach to the problem of systematic sampling was formulated, and the associated variance obtained. Several forms of the population were assumed. The efficiency of the systematic design as compared with the random and stratified random design was evaluated for these forms. It was remarked that as the size of sample increased the variance of a systematic design might also increase, contrary to the behavior of variances in the random sampling design. This possibility was verified in [2].

One approach to the study of systematic designs, given by Cochran [3] removed this difficulty to some extent by changing the problem to one of the expected variance, and supposing the elements of the population to be random variables. He showed that if the correlogram of these random variables is concave upwards, then the expected variance of the systematic design would be less, and often considerably less, than the variance of a stratified design.

In the present paper the results of the earlier papers are extended to the systematic sampling of clusters of equal and unequal sizes. Some comments on systematic sampling in two dimensions are included.

In section 2 we derive two theorems that have considerable applications in many parts of sampling. Although it has been common for people working in sampling theory to tell each other that these theorems ought to be true, yet no reference seems to exist.

In section 3 we develop the implications of a remark [1, p. 13] that in designing sample surveys we should try to induce negative correlation between strata. In Theorem 3 we obtain sufficient conditions for the correlation to be negative. The lemma and Theorem 4 given in Section 4 enable us to extend the uses of Theorem 3 in practice. As an application of these results, we show that if a population has a concave upwards correlogram, and if strata are defined in an optimum fashion for the selection of one element at random from each stratum, then we can define a systematic type design that will be more efficient than independent random selection from each stratum.

In sections 5 and 6 we obtain various results in the systematic sampling of clusters largely as applications of the more general theorems of the earlier sections. In general the results are of a nature similar to those of [1] and [3] in that the formulae show the conditions under which systematic sampling may be expected to be more efficient than random or stratified random sampling. We have not, however, applied these formulae to specified types of populations.

---

[1] Submitted for publication, November, 1948. Parts of this paper were prepared while the author was Visiting Professor of Statistics at the University of São Paulo, Brazil.

[2] References to the articles and book cited are given by Roman numerals.

From [1, 2 and 3] it is already apparent that this work will be useful and such studies should be more valuable when made in connection with important types of surveys or data than when made as illustrations in a general paper.

2. Random events and conditional expectations. Almost invariably, samples are selected in several stages. For example, to select a sample of households from a city one frequently used method is the following two stage sampling plan:

a. A map of the city showing the location of each block is obtained and brought up-to-date.
b. Using this map, a sample of the blocks of the city is selected (this is stage1).
c. From the households on the blocks selected in stage 1, a subsample of households is selected (this is stage 2.).

In this section, we give a general approach for evaluating the means and variances associated with multi-stage sampling. This approach has the advantage of at once yielding the contributions to the variance arising from each stage. Furthermore, the theorems presented are useful in calculating variances even when our interest is not in multi-stage sampling. The theorems are presented in general terms because of their wide application in sampling.

We shall say that the result of performing an operation is a random event A^* if the result can assume m possible states $A_1, \cdots, A_m$ with probabilities $p_1, \cdots, p_m$, where

$$P\{A^* = A_i\} = p_i, \qquad \sum_{i=1}^{m} p_i = 1,$$

and $P\{A^* = A_i\}$ is read "the probability that the random event A^* assumes the state A_i."

One illustration of an operation is the operation of selecting a sample of blocks. If there are N blocks in the city of which we select n in such a way that each set of n of the N blocks is a possible sample, then there are C_n^N possible samples. In this case $m = C_n^N$ and the C_n^N possible samples are the m states of A^* "the result of selecting the sample of blocks." Furthermore, if each of the possible samples of blocks is equally likely to be selected, then

$$P\{A^* = A_i\} = \frac{1}{C_n^N} = \frac{1}{m}.$$

The random event A^* may also be the taking on by a random variable of one of its possible values. If z^* is a random variable having possible values $z_1, \cdots, z_m$ with probabilities $p_1, \cdots, p_m$ then we can define the states of A^* to be A_i where A_i is "$z^* = z_i$."

Thus the notion of a random event includes the two types of randomness that are met in selecting samples.

Let x' be a random variable. Then, by the conditional expectation of x' subject to the random event A^* is meant the random variable $E^*(x' \mid A)$ whose possible values are $E(x' \mid A_i)$, $i = 1, \cdots, m$ and whose probabilities are p_i, that is

$$P\{E^*(x' \mid A) = E(x' \mid A_i)\} = p_i = P\{A^* = A_i\},$$

where

$$E(x' \mid A_i) = \sum_{j=1}^{N_i} x_{ij} p_j(A_i), \tag{2.1}$$

x_{ij} is the jth of the N_i possible values of x' when A_i occurs, and

$$p_j(A_i) = P\{x' = x_{ij} \mid A_i\}$$

is "the probability that $x' = x_{ij}$ given that A_i occurs." It should be noted that if

$$p_{ij} = P\{x' = x_{ij}\},$$

then

$$p_{ij} = P\{x' = x_{ij}, A^* = A_i\}$$

since the fact that $x' = x_{ij}$ implies the occurrence of A_i. Then

$$p_i \cdot p_j(A_i) = p_{ij}. \tag{2.2}$$

We state Theorems 1 and 2 without proof since their proofs are immediate.

THEOREM 1. *The expected value of the random variable $E^*(x' \mid A)$ is $E\,x'$*, i.e.

$$E\{E^*(x' \mid A)\} = Ex'.$$

By $\overset{*}{\sigma}_{x'y'|A}$ we shall mean the random variable whose possible values are $\sigma_{x'y'|A_i}$, $i = 1, \cdots, m$ where

$$\sigma_{x'y'|A_i} = E\{[x' - E(x' \mid A_i)]\,[y' - E(y' \mid A_i)] \mid A_i\}$$

and

$$P\{\sigma_{x'y'|A} = \sigma_{x'y'|A'}\} = p_i = P\{A^* = A_i\},$$

i.e.

$$\overset{*}{\sigma}_{x'y'|A} = E^*\{[x' - E^*(x' \mid A)]\,[y' - E^*(y' \mid A)] \mid A\}.$$

Furthermore, the symbol $\sigma_{E^*(x'|A)E^*(y'|A)}$ will stand for "the covariance of the two random variables $E^*(x' \mid A)$ and $E^*(y' \mid A)$." The corresponding definitions of variance are obtained by replacing y' by x' above.

THEOREM 2. *If x' and y' are random variables, then*

$$\sigma_{x'y'} = E\overset{*}{\sigma}_{x'y'|A} + \sigma_{E^*(x'|A)\,E^*(y'|A)}$$

and

$$\sigma^2_{x'} = E\sigma^2_{x'|A} + \sigma^2_{E^*(x'|A)}.$$

We note that, since the p_{ij}, p_i and $p_j(A_i)$ are not specified, Theorems 1 and 2 are valid for any two-stage plan. The generalizations of Theorems 1 and 2 to multi-stage plans are obvious, but in practice it often turns out to be simpler to apply the theorems several times.

It would be easy to give applications of Theorems 1 and 2 but these are not essential for our purposes in this paper. As remarked in the introduction, these two theorems have long been part of what we may call the folklore of sampling.

3. Stratified sampling and negative correlation, with an application to systematic sampling. In discussing plans for sampling from a stratified population it is customary to suppose that if x' is an estimate and $x' = x_1' + \cdots + x_L'$ where x_j' is the contribution to x' arising from the jth of the L strata, then the sampling is to be so done that the random variables x_i' and x_j', $j \neq i$, are independent.

In [1, p. 13] it was noted that if a population were stratified, and if the elements were so selected that the contributions from different strata were negatively correlated, it would follow that the variance of the estimate would be less than if the contributions were independent but had the same covariances within strata. This was, of course, an immediate conclusion from the fact that

$$\sigma_{x'}^2 = \sum_{i,j=1}^{L} \sigma_{x_i' x_j'}$$

and, hence, if

$$C = \sum_{i \neq j} \sigma_{x_i' x_j'} < 0 \tag{3.1}$$

then $\sigma_{x'}^2$ is less than it would be if $C = 0$. If $C < 0$ we shall say that the sample design has "negative correlation."

It is obvious that any population may be taken to be itself a sample, a sample from the possible populations that might have been produced by the forces that determined the existing population. Inasmuch as sampling designs are often chosen on the basis of a knowledge of the dominating forces and some past experience, it is realistic to consider not only the expected values and variances for a specific population but also their expected values over all possible populations determined by the same forces. Cochran [3] has given one illustration of the usefulness of considering the expected variance of a sample design. He considered the elements $x_1, \cdots, x_n$ of the population themselves to be random variables and supposed that $E\, x_i = \mu$ and $E(x_i - \mu)^2 = \sigma^2$. For his purposes it was also convenient to suppose that if $u > 0$ then $E(x_i - \mu)(x_{i+u} - \mu) = \rho_u \sigma^2$. It was then possible for him to make realistic hypotheses concerning the correlogram, i.e. the ρ_u considered as a function of u, that would not have been reasonable in dealing with a specific population. He thus obtained general conclusions concerning the expected efficiency of systematic sampling designs as compared with random and stratified random designs.

In this paper we shall consider not only the expected values and variances for the given finite population but also the expected values of these expected values and variances under the assumption that the elements of the population are themselves random variables. We shall use $\mathcal{E}$ to denote the expected value considering the elements of the population to be random variables and as before use E for expected values based on the specified finite population.

Then

$$\mathcal{E}\sigma_{x'}^2 = \sum_{i,j=1}^{L} \mathcal{E}\sigma_{x_i' x_j'},$$

and if $\mathcal{E}C < 0$ we shall say that the design has 'expected negative correlation.'

We now propose to obtain the beginnings of an approach to sample design when it is possible to introduce or take advantage of negative correlation or expected negative correlation through the sample design.

To simplify, we shall begin by considering two strata and shall suppose that the possible values of x' are $x_1, \cdots, x_n$ while the possible values of y' are $y_1, \cdots, y_n$. Furthermore, we shall suppose the sampling to be so done that

$$P\{x' = x_i\} = P\{y' = y_i\} = P\{x' = x_i, y' = y_i\} = p_i > 0,$$

so that $\sum_{i=1}^{n} p_i = 1$ and $P\{x' = x_i, y' = y_j\} = 0$ if $i \neq j$.

Under the above assumptions, it follows that

$$\sigma_{x'y'} = \sum_{i=1}^{n} p_i x_i y_i - \sum_{i,j=1}^{n} p_i p_j x_i y_j. \tag{3.2}$$

The symbol $\varphi_{ij} \dot{\geq} 0$ means that $\varphi_{ij} \geq 0$ for all i and j and $\varphi_{ij} > 0$ for at least one pair i, j. We shall say that if $(x_i - x_j)(y_i - y_j) \dot{\geq} 0$ then the sets (x) and (y), where (x) stands for $x_1, \cdots, x_n$ and (y) for $y_1, \cdots, y_n$ are similarly ordered and if $(x_i - x_j)(y_i - y_j) \dot{\leq} 0$ then these sets are oppositely ordered. Then it is easy to prove, [4, p. 43] directly that if the values are oppositely ordered, then $\sigma_{x'y'} < 0$ and if they are similarly ordered then $\sigma_{x'y'} > 0$.

A somewhat more general result is the following:

THEOREM 3. *Let $n \leq k$, let*

$$b = \sum_{i=1}^{n} \sum_{j=1}^{k} a_{ij} w_i z_j$$

be a real bilinear form, and let

$$t = \sum_{i=j}^{n} a_{ii} w_i$$

be a real linear form, where $w_i > 0$, $z_i > 0$ and $\sum_{i=1}^{n} w_i = \sum_{i=1}^{k} z_i = 1$.

Then a sufficient condition that $b > t$ is

$$a_{ij} \dot{\geq} a_{ij}. \tag{3.3}$$

If $k = n$ and $w_i = z_i$ then $b > t$ if

$$a_{ij} + a_{ji} \dot{\geq} a_{ii} + a_{jj}. \tag{3.4}$$

PROOF. Since

$$b - t = \sum_{i=1}^{n} a_{ii}(w_i z_i - w_i) + \sum_{i \neq j} a_{ij} w_i z_j ,$$

and since

$$1 - z_i = \sum_{\substack{j=1 \\ j \neq i}}^{k} z_j ,$$

it follows that

$$b - t \sum_{j \neq i} (a_{ij} - a_{ii}) w_i z_j .$$

Hence, $b > t$ if (3.3) holds. Also, if $k = n$ and $w_i = z_i$ then $b > t$ if (3.4) holds. Some obvious generalizations of Theorem 3 have been omitted since we do not need them.

To obtain the result that $\sigma_{x'y'} < 0$ if the sets (x) and (y) are oppositely ordered, we make the identifications $a_{ij} = x_i y_j$ and $z_i = w_i = p_i$. Then (3.4) holds and substituting we have

$$a_{ii} + a_{jj} - a_{ij} - a_{ji} = (x_i - x_j)(y_i - y_j) \tag{3.5}$$

so that if the values are oppositely ordered, $\sigma_{x'y'} < 0$, and hence the two strata have negative correlation.

To consider expected negative correlation we note that

$$\mathcal{E}\sigma_{x'y'} = \sum_{i=1}^{n} p_i \sigma_{ii} + \sum_{i,j=1}^{n} p_i p_j \sigma_{ij} \tag{3.6}$$

where we suppose that $\mathcal{E}x_i = \mu$, $\mathcal{E}y_i = \nu$ and

$$\mathcal{E}(x_i - \mu)(y_j - \nu) = \sigma_{ij}$$

so that in this case σ_{ii} is a covariance, not a variance.

If we put $a_{ij} = \sigma_{ij}$ and $z_i = w_i = p_i$, then (3.4) holds and we obtain, as sufficient for $\mathcal{E}\sigma_{x'y'}$ to be negative, that

$$\sigma_{ij} + \sigma_{ji} \geq \sigma_{ii} + \sigma_{jj} \tag{3.7}$$

or, if we define ρ_{ij} by the equation,

$$\sigma_x \sigma_y \rho_{ij} = \sigma_{ij} ,$$

where $\sigma_x^2 = \mathcal{E}(x_i - \mu)^2$ and $\sigma_y^2 = \mathcal{E}(y_i - \nu)^2$, we have

$$\rho_{ij} + \rho_{ji} \geq \rho_{ii} + \rho_{jj} \tag{3.8}$$

as a sufficient condition for $\mathcal{E}\sigma_{x'y'} < 0$.

Let us consider the systematic sampling of single elements. In systematic sampling, we assume a population of kn ordered elements $x_1, x_2, \cdots, x_k, x_{1+k}, \cdots, x_{2k}, \cdots, x_{1+(n-1)k}, \cdots, x_{nk}$ of which we wish to estimate the arithmetic mean $\bar{x}$. As our estimate we use

$$\bar{x}' = (x_1' + \cdots + x_m')/n$$

where x_1' is selected at random from $x_1, \cdots, x_k$ and if $x_1' = x_j$ then $x_i' = x_{j+(i-1)k}$, $i = 2, \cdots, m$. Thus, $\bar{x}'$ may be interpreted as an estimate based on a stratified population, the ith stratum consisting of

$$x_{1+(i-1)k}, \cdots, x_{k+(i-1)k}$$

and

$$P\{x_i' = x_{\alpha+(i-1)k}\} = P\{x_i' = x_{\alpha+(i-1)k}, x_j' = x_{\alpha+(j-1)k}\} = 1/k$$

while

$$P\{x_i' = x_{\alpha+(i-1)k}, x_j' = x_{\beta+(j-1)k}\} = 0, \text{ if } \alpha \neq \beta.$$

Then

$$\sigma_{x_i'x_j'} = \left(\frac{1}{k}\right) \sum_{\alpha=1}^{k} x_{\alpha+(i-1)k} \cdot x_{\alpha+(j-1)k} - \bar{x}_i \bar{x}_j$$

where

$$\bar{x}_i = \left(\frac{1}{k}\right) \sum_{\alpha=1}^{k} x_{\alpha+(i-1)k}.$$

Hence, any two strata that are oppositely ordered will yield a negative contribution to the variance. However, since it is not possible for all strata to be negatively ordered, we do not thus obtain a useful result and must return to the consideration of C or $\sigma_{\bar{x}'}^2$ itself as was done in [1]. If, however, we make Cochran's assumptions, and consider $\mathcal{E}\sigma_{x'y'}$, it follows that for the ith and jth strata

$$\rho_{\alpha\beta} = \rho_{(j-i)k+\beta-\alpha},$$

and (3.8) becomes

$$\rho_{(j-i)k+(\beta-\alpha)} + \rho_{(j-i)k+(\alpha-\beta)} \geq 2\rho_{(j-i)k}, \tag{3.9}$$

i.e. the correlation function ρ_u must be concave upwards, which Cochran showed by other means. By considering $\mathcal{E}C$ it is possible to show that a sort of average concavity is all that is required of the correlogram for systematic sampling to have a smaller variance than stratified random sampling.

4. Conditions for negative correlation when the strata are of unequal sizes with an application to systematic sampling. Often, as in the systematic selection of clusters with probability proportionate to size (discussed in Section 5) the simplified situation dealt with in Theorem 3 does not directly apply. However, Theorem 3 may be used to advantage by the following device.

Let us suppose the possible values of x' to be $x_1, \cdots, x_n$ and those of y_0' to be $y_1^0, \cdots, y_k^0$, $k > n$ and let

$$P\{y' = y_\beta^0 \mid x' = x_\alpha\} = p_{\beta|\alpha}$$

so that if we define

$$y_\alpha = \sum_{\beta=1}^{k} y_\beta^0 p_{\beta|\alpha}, \tag{4.1}$$

then

$$y_\alpha = E(y_0' \mid x' = x_\alpha).$$

If we define y' to be a random variable having possible values $y_1, \cdots, y_n$ with probabilities $p_1, \cdots, p_n$ where

$$p_\alpha = P\{x' = x_\alpha\}$$

it follows that

$$y' = E^*(y_0' \mid x')$$

and

$$\sigma_{x'y_0'} = \sigma_{x'y'}.$$

Clearly, Theorem 3 is valid for the random variables x' and y'.

Consequently, we need only determine what restrictions the conditional probabilities, $p_{\beta|\alpha}$, and the values, y_α^0, need satisfy for the sets $x_1, \cdots, x_n$ and $y_1, \cdots, y_n$ to be oppositely ordered or for (3.7) to hold.

Substituting for y_i and y_j in (3.5) we see that if

$$(x_\alpha - x_\gamma) \sum_{\beta=1}^{k} y_\beta^0 (p_{\beta|\alpha} - p_{\beta|\gamma}) \leq 0 \tag{4.2}$$

then $\sigma_{x'y'} = \sigma_{x'y_0'} < 0$.

Let

$$\sigma_{\alpha\gamma}^0 = \mathcal{E}(x_\alpha - \mu)\,(y_\gamma^0 - \nu).$$

Then substituting in (3.7) we see that if

$$\sum_{\beta=1}^{k} (p_{\beta|\alpha} - p_{\beta|\gamma})(\sigma_{\alpha\beta}^0 - \sigma_{\gamma\beta}^0) \mathrel{\dot{\leq}} 0 \tag{4.3}$$

or if

$$\sum_{\beta=1}^{k} (p_{\beta|\alpha} - p_{\beta|\gamma})(\rho_{\alpha\beta}^0 - \rho_{\gamma\beta}^0) \mathrel{\dot{\leq}} 0 \tag{4.4}$$

then

$$\mathcal{E}\sigma_{x'y_0'} < 0.$$

In order to use (4.2) and (4.3) the following well-known lemma is often useful.

Lemma. *If $\xi_1 \leq \xi_2 \leq \cdots \leq \xi_k \leq 0$ and the quantities $\epsilon_1, \cdots, \epsilon_k$ are such that*

$$\sum_{\beta=1}^{s} \epsilon_\beta \geq 0$$

then

$$\sum_{\beta=1}^{s} \epsilon_\beta \xi_\beta \leq 0, \qquad s = 1, \cdots, k.$$

Let us use this lemma to obtain another theorem that will be helpful in showing negative or expected negative correlation between strata.

THEOREM 4. *Let b be a bilinear form*

$$b = \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} w_i z_j$$

such that $\sum_{i=1}^{s} w_i \geq 0$, $\sum_{j=1}^{s'} z_j \geq 0$, $s = 1, \cdots, n-1$, $s' = 1, \cdots, m-1$, *and*

$$\sum_{i=1}^{n} w_i = \sum_{j=1}^{m} z_j = 0. \tag{4.5}$$

Let

$$\delta_{ij} = a_{ij} - a_{i+1,j} - a_{i,j+1} + a_{i+1,j+1}.$$

Then a sufficient condition that $b \leq 0$ is $\delta_{ij} \leq 0$.

PROOF. Upon substituting for w_n and z_m in b from (4.5) we see that

$$b = \sum_{i=1}^{n-1} \sum_{j=1}^{m-1} \delta'_{ij} w_i z_j$$

where

$$\delta'_{ij} = a_{ij} - a_{im} - a_{nj} + a_{nm}$$

or, if we define,

$$\xi_j = \sum_{i=1}^{n-1} \delta'_{ij} w_i$$

then

$$b = \sum_{j=1}^{m-1} \xi_j z_j.$$

According to the lemma, it then follows that a sufficient condition that $b \leq 0$ is that

$$\xi_1 \leq \xi_2 \leq \cdots \leq \xi_{m-1} \leq 0.$$

Also, a sufficient condition that

$$\xi_j - \xi_{j+1} \leq 0,$$

is

$$\delta'_{ij} - \delta'_{i,j+1} \leq \delta'_{i+1,j} - \delta'_{i+1,j+1}$$

Then to complete the proof it is only necessary to verify that

$$\delta_{ij} = \delta'_{ij} - \delta'_{i,j+1} - \delta'_{i+1,j} + \delta'_{i+1,j+1} .$$

In the preceding pages we have given an identification of systematic with stratified sampling where, instead of the selection being made independently within strata, the choice of an element from one stratum determines the choice from the other strata. In this identification, however, it was assumed that the strata contained the same number of elements. Let us now extend this method of selecting samples to the case where the strata have different numbers of elements. In so doing we shall illustrate the use of the above lemma and theorem 4.

Suppose now that the population consists of N elements $x_1, \cdots, x_N$ classified into n strata, the ith of which contains the N_i elements

$$x_{N_1+\cdots+N_{i-1}+1}, \cdots, x_{N_1+\cdots+N_i} .$$

We shall denote these elements by $x_{i1}, \cdots, x_{iN_i}$.

We shall select one element from each of these n strata. The element selected from the ith stratum is written x'_i. As the estimate of $\bar{x}$, the arithmetic mean of the population, we use

$$\bar{x}' = \sum_{i=1}^{n} \frac{N_i}{N} x'_i$$

and it is well known that if the selection is made independently at random from each stratum, then

$$\sigma^2_{\bar{x}'} = \sum_{i=1}^{n} \left(\frac{N_i}{N}\right)^2 \sigma^2_i$$

where σ^2_i is the variance of x'_1, i.e. the variance of the ith stratum.

Let us now consider an alternative to the usual method. We can suppose that $N_1 > 1$ without any loss of generality. (The methods are the same for any stratum having $N_i = 1$ and will also yield the same result for any population such that either all the $N_i = 1$ or all but one of the $N_i = 1$. Differences occur if at least two of the N_i differ from 1.)

We first choose an element at random from the first stratum. Suppose that $x'_1 = x_\alpha$. Then to choose an element from the second stratum, assuming that $N_2 > 1$, we proceed as follows: Multiply N_2 by any positive integer t_2 such that N_2t_2/N_1 is an integer, say, k_2. Assign to each element of the second stratum the measure of size t_2, and form the two sets of cumulative totals $t_2, 2t_2, \cdots, N_2t_2$ and $k_2, 2k_2, \cdots, N_1k_2$. Then with the measures of size t_2 assigned to each element of stratum 2, and the measure of size k_2 assigned to each element of stratum 1, it follows that strata 1 and 2 have the same total size.

As an example of the arithmetic given below consider the following simple case. Suppose that $N_1 = 3$ and $N_2 = 4$. Then if we take for t_2 the value 6, it follows that $k_2 = 8$. We choose one of the integers 1, 2, 3 with equal probability. If the integer 1 is obtained, we have selected the first element of the first stratum and choose an integer between 1 and 8 with equal probability. If the selected integer is between 1 and 6, the first element of the second stratum is selected. If it is 7 or 8 the second element of the second stratum is selected. Similarly if the second element of the first stratum is selected, then we select an integer between 9 and 16 with equal probability. If that integer has value 9, $\cdots$, 12 the second element of the second stratum is selected; if it has value 13, $\cdots$, 16 the third element is selected.

The general formulation of the selection procedure for the second stratum is:

Suppose that β_0 is the smallest integer such that $(\alpha - 1)k_2 + 1 \leq \beta_0 t_2$ and that β_1 is such that $(\beta_1 - 1)t_2 < \alpha k_2 \leq \beta_1 t_2$. Choosen an integer at random from $1, \cdots, k_2$ and call that integer β. Then, if

$$(\alpha - 1)k_2 < (\alpha - 1)k_2 + \beta \leq \beta_0 t_2$$

the β_0th element is selected from stratum 2; if

$$\beta_0 t_2 < (\alpha - 1)k_2 + \beta \leq (\beta_0 + 1)t_2$$

the $(\beta_0 + 1)$th element is selected; $\cdots$; and if

$$(\beta_1 - 1)t_2 < (\alpha - 1)k_2 + \beta \leq \alpha k_2$$

the β_1th element is selected from stratum 2.

It is easy to verify that when the sample is so selected, each element of stratum 2 has equal probability of being selected. Hence, if we apply this procedure to each stratum we have

$$\sigma^2_{\bar{x}'} = \sum_{i=1}^{n} \left(\frac{N_i}{N}\right)^2 \sigma_i^2 + \sum_{i \neq j} \frac{N_i N_j}{N^2} \sigma_{x_i' x_j'}.$$

Let us evaluate $\sigma_{x_i' x_j'}$ for this type of selection. Now

$$\sigma_{x_i' x_j'} = E\,(x_i' - \bar{x}_i)(x_j' - \bar{x}_j)$$

where $\bar{x}_i$ is the arithmetic mean of the elements of the ith stratum. From Theorem 2, we then have

$$\sigma_{x_i' x_j'} = \frac{1}{N_1} \sum_{\alpha=1}^{N_1} E[x_i' - E(x_i' \mid x_{1\alpha})][x_j' - E(x_j' \mid x_{1\alpha}]$$
$$+ \frac{1}{N_1} \sum_{\alpha=1}^{N_1} [E(x_i' \mid x_{1\alpha}) - \bar{x}_i][E(x_j' \mid x_{1\alpha}) - \bar{x}_j].$$

It is easy to see that the method of selection used above implies that the first term of $\sigma_{x_i' x_j'}$ vanishes. Furthermore, $\bar{x}_i$ is the arithmetic mean of the conditional expectations so that we have reduced the problem to one of determining whether the conditional expectations satisfy the conditions for negative correlation or expected negative correlation.

If we denote $E\,(x_i' \mid x_{1\alpha})$ by $y_{i\alpha}$, then we need to see whether the sets $y_{i1} \cdots$,

y_{iN_1} and $y_{ji}, \cdots, y_{jN_1}$ are oppositely ordered. Now

$$(y_{i\alpha} - y_{i\beta})(y_{j\alpha} - y_{j\beta} = \sum_{g=1}^{N_i} \sum_{h=1}^{N_j} x_{ig} x_{jh} \epsilon_{ig\alpha\beta} \epsilon_{jh\alpha\beta}$$

where

$$\epsilon_{ig\alpha\beta} = P\{x_i' = x_{ig} \mid x_{1\alpha}\} - P\{x_i' = x_{ig} \mid x_{1\beta}\}.$$

If $\alpha < \beta$ then, according to the method of selection,

$$\sum_{g=1}^{s} \epsilon_{ig\alpha\beta} \geq 0, \qquad s = 1, \cdots, N_i - 1$$

while

$$\sum_{g=1}^{N_i} \epsilon_{ig\alpha\beta} = 0.$$

In Theorem 4, we then make the identifications $n = N_i$, $m = N_j$,

$$w_g = \epsilon_{ig\alpha\beta}, \qquad z_h = \epsilon_{jh\alpha\beta} \quad \text{and} \quad a_{gh} = x_{ig} x_{jh}.$$

Then

$$\delta_{gh} = (x_{ig} - x_{i,g+1})(x_{jh} - x_{j,h+1})$$

and hence to have negative correlation between the strata, it is sufficient that the sets $x_{i1}, \cdots, x_{iN_i}$ and $x_{ji}, \cdots, x_{jN_j}$ have the type of negative ordering represented by $\delta_{gh} \leq 0$. Similarly, if

$$\sigma_{gh} = \mathcal{E}(x_{ig} - \mu_i)(x_{jh} - \mu_j), \qquad \mu_i = \mathcal{E}x_{ig},$$

then, for expected negative correlation, it is sufficient that

$$\sigma_{gh} - \sigma_{g,h+1} - \sigma_{g+1,h} + \sigma_{g+1,h+1} \leq 0.$$

Of course, these conditions will be satisfied if a concave upwards correlogram exists. Hence, if a population consists of N random variables $x_1, \cdots, x_N$ having a concave upwards correlogram, then, no matter into what strata these elements are classified, provided that the order of occurrence of the elements remains unaltered, the systematic selection of the elements in the sample can be so planned as to yield an estimate having smaller variance than the stratified random selection of the elements in the sample even if optimum allocation is used. If more than one element is being selected from a stratum under optimum allocation, then the systematic selection of the same number of elements will suffice. If not only optimum allocation but also optimum definitions of strata are being used so that but one element is selected from each stratum, then systematic selection according to the scheme described in the section will produce a variance not larger than the variance of stratified random sampling. It should be noted, however, that this does not imply that a 'hammer and tongs' use of systematic sampling ignoring the strata will produce a smaller variance. There is work to be done on what is required for the latter to occur.

It may be noted that the procedure of this example provides an answer to the systematic selection of elements from a population whose size is not a multiple of the size of sample.

5. The systematic sampling of clusters with probability proportionate to a measure of size. It is known [5] that sampling clusters with probability proportionate to a measure of size often yields considerable reductions in the variance of the estimates. However, the theory of the systematic selection of several clusters with probability proportionate to a measure of size has not been worked out, and it is the purpose of this section to make some contributions to that theory.

The most frequently used method of sampling clusters with probability proportionate to size is equivalent to the following: Suppose that the clusters are denoted by $C_1, \cdots, C_M$ and that to the hth of these M clusters is assigned a measure of size P_h. Form the successive totals $P_1, P_1 + P_2, P_1 + P_2 + P_3, \cdots, P_1 + \cdots + P_M$. If we wish to select m of these clusters, we calculate $\bar{P}_m = (P_1 + \cdots + P_M)/_m$. Then, assuming that $P_j \leq \bar{P}_m$, $j = 1, \cdots, M$, we select an integer with equal probability from $1, \cdots, \bar{P}_m$. Calling that integer P', we calculate the m numbers $P', P' + \bar{P}_m, P' + 2\bar{P}_m, \cdots, P' + (m - 1)\bar{P}_m$. If

$$(5.1) \qquad P_1 + \cdots + P_{h-1} + 1 \leq P' + (i - 1)\bar{P}_m \leq P_1 + \cdots + P_h$$

for any integer i, $i = 1, \cdots, m$, then the cluster C_h is selected for the sample. Any cluster for which $P_h > \bar{P}_m$ is automatically included in the sample, and if there are, say, α such clusters, then we calculate $\bar{P}_{m-\alpha}$ for the $M - \alpha$ clusters remaining after including these α in the sample, and proceed as above.

In deriving the variance of the estimate we shall use, we interpret that estimate as a stratified sampling estimate. Although it is easy to obtain the expected value of the estimate without that interpretation, we shall need it later in the derivation of the variance, and hence we give it here to shorten the total presentation a little.

Suppose that clusters $C_1, \cdots, C_{k_1}$ are such that

$$P_1 + \cdots + P_{k_1-1} < \bar{P}_m \leq P_1 + \cdots + P_{k_1}.$$

Then we define stratum 1 to consist of clusters $C_1, \cdots, C_{k_1}$. It is easy to see that if the above sampling method is used then

$$P\{C_h \text{ is selected from stratum } 1, h < k_1\} = \frac{P_h}{\bar{P}_m},$$

$$P\{C_{k_1} \text{ is selected from stratum } 1\} = \frac{\bar{P}_m - P_1 - \cdots - P_{k_1-1}}{\bar{P}_m}.$$

Furthermore, suppose that clusters $C_{k_1}, \cdots, C_{k_1+k_2}$ are such that

$$P_1 + \cdots + P_{k_1+k_2-1} < 2\,\bar{P}_m \leq P_1 + \cdots + P_{k_1+k_2}.$$

Then we define stratum 2 to consist of clusters $C_{k_1}, \cdots, C_{k_1+k_2}$. It is easy see that if the above sampling method is used, then

$$P\{C_{k_1} \text{ is selected from stratum 2}\} = \frac{P_1 + \cdots + P_{k_1} - \bar{P}_m}{\bar{P}_m},$$

$$P\{C_{k_1+h} \text{ is selected from stratum 2}, 1 \leq h < k_2\} = \frac{P_h}{\bar{P}_m},$$

$$P\{C_{k_1+k_2} \text{ is selected from stratum 2}\} = \frac{2\bar{P}_m - P_1 - \cdots - P_{k_1+k_2-1}}{\bar{P}_m}.$$

Since $P_h \leq \bar{P}_m$ we remark that it is impossible that C_{k_1} be selected from both stratum 1 and stratum 2.

In general, if clusters $C_{k_1+\cdots+k_{i-1}}, \cdots, C_{k_1+\cdots+k_i}$ are such that

$$P_1 + \cdots + P_{k_1+\cdots+k_{i-1}} < i\bar{P}_m \leq P_1 + \cdots + P_{k_1+\cdots+k_i} \tag{5.2}$$

then the ith stratum consists of these $k_i + 1$ clusters, and we define the probabilities $P_{i\alpha}$, $\alpha = 0, \cdots, k_i$, by the equations

$$\begin{aligned} P_{i0} &= P\{C_{k_1+\cdots+k_{i-1}} \text{ is selected from stratum } i\} \\ &= \frac{P_1 + \cdots + P_{k_1+\cdots+k_{i-1}} - (i-1)\bar{P}_m}{\bar{P}_m}, \\ P_{i\alpha} &= P\{C_h \text{ is selected from stratum } i, k_1 + \cdots + k_{i-1} < h < k_1 + \cdots + k_i\} \\ &= \frac{P_h}{\bar{P}_m}, \qquad \alpha = h - k_1 - \cdots - k_{i-1}, \\ P_{ik_i} &= P\{C_{k_1+\cdots+k_i} \text{ is selected from stratum } i\} \\ &= \frac{i\bar{P}_m - P_1 - \cdots - P_{k_1+\cdots+k_{i-1}}}{\bar{P}_m}. \end{aligned} \tag{5.3}$$

We remark that

$$P_{i-1k_{i-1}} + P_{i0} = \frac{P_{k_1+\cdots+k_{i-1}}}{\bar{P}_m}. \tag{5.4}$$

Now, let the elements of the population be x_{hj}, $h = 1, \cdots, M$, $j = 1, \cdots, N_h$, and let the arithmetic mean of the hth cluster be denoted by $\bar{x}_h$. Since the N_h are usually unknown but the measure of size, P_h, is known, we sample, not with probability proportionate to the N_h, but with probability proportionate to the P_h. We shall denote the clusters of the ith stratum by $C_{i0}, \cdots, C_{ik_i}$, making the identification

$$C_{i\alpha} = C_{\alpha+k_1,+\cdots+k_{i-1}}. \tag{5.5}$$

Furthermore, the number of elements of the clusters are denoted by $N_{i0}, \cdots,$

N_{ik_i}, and the means of the clusters by $\bar{x}_{i0}, \cdots, \bar{x}_{ik_i}$, where

$$N_{i\alpha} = N_{\alpha+k_1+\cdots+k_{i-1}}$$

$$\bar{x}_{i\alpha} = \bar{x}_{\alpha+k_1+\cdots+k_{i-1}} \tag{5.6}$$

so that $\bar{x}_{i0} = \bar{x}_{i-1,k_{i-1}}$ and $N_{i0} = N_{i-1,k_{i-1}}$, $i = 1, \cdots, m$.

Furthermore, we define

$$\tilde{x}_{i\alpha} = N_{i\alpha}\,\bar{x}_{i\alpha}/P_{i\alpha} = \tilde{x}_{\alpha+k_1+\cdots+k_{i-1}} \tag{5.7}$$

We define the mean of the ith stratum to be

$$\bar{\tilde{x}}_i = \sum_{\alpha=0}^{k_i} P_{i\alpha}\tilde{x}_{i\alpha}/\bar{P}_m, \tag{5.8}$$

and the variance of the ith stratum to be

$$\sigma_i^2 = \sum_{\alpha=0}^{k_i} \frac{P_{i\alpha}}{\bar{P}_m}(\tilde{x}_{i\alpha} - \bar{\tilde{x}}_i)^2. \tag{5.9}$$

Then, if the mean and variance of the population are defined to be

$$\tilde{x} = \sum_{h=1}^{M} P_h\tilde{x}_h/P \tag{5.10}$$

and

$$\sigma^2 = \sum_{h=1}^{M} \frac{P_h}{P}(\tilde{x}_h - \tilde{x})^2, \tag{5.11}$$

it is easy to verify that

$$\tilde{x} = \frac{1}{m}\sum_{i=1}^{m} \bar{\tilde{x}}_i \tag{5.12}$$

and

$$\sigma^2 = \frac{1}{m}\sum_{i=1}^{m}\sigma_i^2 + \frac{1}{m}\sum_{i=1}^{m}(\bar{\tilde{x}}_i - \tilde{x})^2. \tag{5.13}$$

An unbiased estimate of the total of a characteristic. We shall see that we can obtain an estimate of x, where

$$x = \sum_{i=1}^{M}\sum_{j=1}^{N_i} x_{ij}$$

i. e. x is the total of the elements of the population. Since N is unknown, the estimate of $\bar{x}$ that is used is the ratio of unbiased estimates of x and N. It is well known that this ratio is usually biased. Since we are not making any study of ratio estimates here we will not derive the approximation to the variance of this estimate. It may be remarked that it can be obtained by a simple extension of the results here given.

Let us agree that the general form of the estimate will be as follows:

If the jth cluster of the population is selected we shall subsample n_j elements from it. The total of the values of the characteristic for these n_j elements we denote by x_j'. Furthermore, we denote by n_i' the total number of elements subsampled from the ith stratum, or, what is the same, from the cluster selected from the ith stratum; and by x_i'' the total of these elements. Thus, if the jth cluster is the ith selected, then $n_i' = n_j$ and $x_i'' = x_j'$. We define our estimate x'' of x, the total of the population, to be

$$x'' = K(x_1'' + \cdots + x_m''). \tag{5.14}$$

Then, if $K = P/mn$ and $n_h = nN_h/P_h$, it is easy to see that x'' is an unbiased estimate of x.

The variance of the estimate. We may calculate the variance of x'' where

$$x'' = \bar{P}_m(\tilde{x}_1'' + \cdots + \tilde{x}_m'') \quad \text{and} \quad \tilde{x}_i'' = x_i''/n. \tag{5.15}$$

Now, by Theorem 2,

$$\sigma_{x''}^2 = E\sigma_{x''|A}^{2*} + \sigma_{E^*(x''|A)}^2, \tag{5.16}$$

where A^* has been defined above. We shall not evaluate $E\sigma_{x''|A}^{2*}$ since this involves no new problem for subsampling methods using random or systematic methods, or methods using probability proportionate to size.

From (5.15) it follows that

$$E^*(x''|A) = \bar{P}_m(\tilde{x}_1' + \cdots + \tilde{x}_m') \tag{5.17}$$

or, in other words, $E^*(x''|A)$ is the estimate we would have if the clusters in the sample were completely enumerated. We shall denote the second term of (5.16) by σ_B^2. Then,

$$\sigma_B^2 = \bar{P}_m^2\left\{\sum_{i=1}^m \sigma_{\tilde{x}_i'}^2 + \sum_{i\neq j} \sigma_{\tilde{x}_i'\tilde{x}_j'}\right\}. \tag{5.18}$$

Now

$$\sigma_{\tilde{x}_i'}^2 = \sum_{\alpha=0}^{k_i} \frac{P_{i\alpha}}{\bar{P}_m}(\tilde{x}_{i\alpha} - \bar{\bar{x}}_i)^2 = \sigma_i^2. \tag{5.19}$$

To calculate $\sigma_{\tilde{x}_i'\tilde{x}_j'}$, $i \neq j$, we shall use Theorem 1.

$$\sigma_{\tilde{x}_i'\tilde{x}_j'} = E(\tilde{x}_i' - \bar{\bar{x}}_i)(\tilde{x}_j' - \bar{\bar{x}}_j) = E\{(\tilde{x}_i' - \bar{\bar{x}}_i)E^*[(\tilde{x}_j' - \bar{\bar{x}}_j) \mid \tilde{x}_i']\}. \tag{5.20}$$

To calculate $E^*[(\tilde{x}_j' - \bar{\bar{x}}_j) \mid \tilde{x}_i']$ we begin by noting that

$$E^*[(\tilde{x}_j' - \bar{\bar{x}}_j) \mid \tilde{x}_i'] \equiv E^*[(\tilde{x}_j' - \bar{\bar{x}}_j) \mid C_i] \tag{5.21}$$

where C_1^* is the random event having $k_i + 1$ possible states which are the selections of $C_{i0}, \cdots, C_{ik_i}$ as the sample clusters of the ith stratum. Now if $C_{i\alpha}$ is one of the clusters of the ith stratum let us calculate

$$E[(\tilde{x}_j' - \bar{\bar{x}}_j) \mid C_{i\alpha}]. \tag{5.22}$$

We begin by determining which of the clusters of the jth stratum are possible sample clusters, if we know that $C_{i\alpha}$ is selected from the ith stratum. Since the sizes of strata i and j are both $\overline{P}_m$ it follows that there exist integers β_0 and β_1 such that

$$P_{j0} + \cdots + P_{j\beta_0 - 1} \leq P_{i1} + \cdots + P_{i,\alpha-1} < P_{j1} + \cdots + P_{j\beta_0},$$

and

$$P_{j0} + \cdots + P_{j\beta_1 - 1} < P_{i1} + \cdots + P_{i\alpha} \leq P_{j1} + \cdots + P_{j\beta_1}.$$

Hence, if we know that $C_{i\alpha}$ has been selected from stratum i, it follows that we must select one of the clusters

$$C_{j\beta_0}, C_{j\beta_0+1}, \cdots, C_{j\beta_1}$$

from stratum j and

$$\begin{aligned} P\{C_{j\beta} \text{ is selected} \mid C_{i\alpha} \text{ is selected}\} &= P'_{j\beta}/P_{i\alpha}, \beta = \beta_0, \beta_0 + 1, \cdots, \beta_1 \\ &= 0, \text{ otherwise,} \end{aligned}$$

where

$$\begin{aligned} P'_{j\beta_0} &= P_{j1} + \cdots + P_{j\beta_0} - P_{i1} - \cdots, P_{i,\alpha-1} \\ P'_{j\beta} &= P_{j\beta}, \beta = \beta_0 + 1, \cdots, B_1 - 1 \\ P'_{j\beta_1} &= P_{i1} + \cdots + P_{i\alpha} - P_{j1} - \cdots - P_{j\beta_1 - 1}, \end{aligned}$$

and

$$\sum_{\beta=\beta_0}^{\beta_1} P'_{j\beta} = P_{i\alpha}$$

Then

$$E[(\tilde{x}'_j - \overline{\overline{\tilde{x}}}_j \mid C_{i\alpha}) = \tilde{x}_{j|\alpha} - \overline{\overline{\tilde{x}}}_j] \tag{5.23}$$

where

$$\tilde{x}_{j|\alpha} = \sum_{\beta=\beta_0}^{\beta_1} \frac{P'_{j\beta}}{P_{i\alpha}} \tilde{x}_{j\beta}. \tag{5.27}$$

Hence, substituting in (5.20), we see that

$$\sigma_{\tilde{x}'_i \tilde{x}'_j} = E(\tilde{x}'_i - \overline{\overline{\tilde{x}}}_i)(\tilde{x}'_{j|i} - \overline{\overline{\tilde{x}}}_j)$$

where $\tilde{x}'_{j|i} = \tilde{x}_{j|\alpha}$ if $C_{i\alpha}$ is selected from stratum i. Then it follows that

$$\sigma_{\tilde{x}'_i \tilde{x}'_j} = \sum_{\alpha=0}^{k_i} \frac{P_{i\alpha}}{\overline{P}_m} (\tilde{x}_{i\alpha} - \overline{\overline{\tilde{x}}}_i)(\tilde{x}_{j|\alpha} - \overline{\overline{\tilde{x}}}_j). \tag{5.25}$$

Obviously, the conditional expectation can be eliminated from (5.25) by using (5.23) but no gain in simplicity or generality thus occurs.

It would be possible to obtain the variances and covariances of the x_i' by listing all possible samples in any special case. To make this general would only require writing the necessary notation.

Substituting in (5.18) we see that

$$\sigma_B^2 = \bar{P}_m^2 \left\{ \sum_{i=1}^{m} \sigma_i^2 + \sum_{i \neq j} \sigma_{\bar{x}_i' \bar{x}_j'} \right\}$$

where $\sigma_{\bar{x}_i' \bar{x}_j}$ is given by (5.25).

It follows that if we use the fact that $\sum_{i=1}^{m} (\bar{\bar{x}}_i - \bar{x}) = 0$, then we have

$$\sigma_B^2 = \bar{P}_m^2 \sum_{i=1}^{m} \sum_{\alpha=0}^{k_i} \frac{P_{i\alpha}}{\bar{P}_m} (\bar{x}_{i\alpha} - \bar{x})^2 + \bar{P}_m^2 \sum_{i \neq j} \sum_{\alpha=0}^{k_i} \frac{P_{i\alpha}}{\bar{P}_m} (\bar{x}_{i\alpha} - \bar{x})(\bar{x}_{j|\alpha} - \bar{x}),$$

or, returning in part to the "unstratified" notation

$$\text{(5.26)} \quad \sigma_B^2 = \frac{P^2}{m} \sum_{h=1}^{M} \frac{P_h}{P} (\bar{x}_h - \bar{x})^2 + \frac{P^2}{m} \sum_{i \neq j} \sum_{\alpha=0}^{k_i} \frac{P_{i\alpha}}{P} (\bar{x}_{i\alpha} - \bar{x})(\bar{x}_{j|\alpha} - \bar{x}).$$

By combining terms of the second part of (4.26) generalizations of the formulae obtained in [1] are easily obtained.

Still another means of writing σ_B^2 is

$$\text{(5.27)} \quad \sigma_B^2 = \frac{P^2}{m} \left\{ \sigma^2 - \sigma_{b \cdot s \cdot}^2 + \sum_{i \neq j} \sum_{\alpha=0}^{k_i} \frac{P_{i\alpha}}{P} (\bar{x}_{i\alpha} - \bar{x})(\bar{x}_{j|\alpha} - \bar{x}) \right\}$$

where

$$\sigma_{b \cdot s \cdot}^2 = \frac{1}{m} \sum_{i=1}^{m} (\bar{\bar{x}}_i - \bar{x})^2,$$

which shows both sources of changes in efficiency as compared with sampling with probability proportionate to size, and replacing the clusters obtained. (It is, of course, obvious that $P^2\sigma^2/m$ is the variance of $E^*(x'' \mid A)$, if we assume the m clusters to have been selected with probability proportionate to size, each selected cluster being replaced before the next is selected.)

By considering (5.26) and (5.27) it is clear that systematic sampling with probability proportionate to size will be more efficient than sampling p.p.s. with replacement under much the same conditions as when we sample single elements. The details are omitted. They depend on applying the Lemma and Theorem 4. The summary of the conditions is: If we sample systematically with p.p.s., and if the two sets $x_{i1}, \cdots, x_{ik_i}$ and $y_{ji}, \cdots y_{jk_j}$ are monotone, one being monotone non-icreasing and the other monotone non-decreasing, then the covariance between the ith and jth strata will be negative, and thus gains made as compared with independent sampling from the strata.

If we define

$$\sigma_{\alpha\beta}^0 = \mathcal{E}(\bar{x}_{i\alpha} - \mathcal{E}\bar{x}_{i\alpha})(\bar{x}_{j\beta} - \mathcal{E}\bar{x}_{j\beta})$$

then the concavity condition for systematic sampling p.p.s. to yield a smaller variance than independent sampling p.p.s. from each stratum is, if $\alpha < \beta$,

$$\sigma^0_{\alpha 1} - \sigma^0_{\gamma 1} \leq \sigma^0_{\alpha 2} - \sigma^0_{\gamma 2} \leq \cdots \leq \sigma^0_{\alpha k_f} - \sigma^0_{\gamma k_f} \leq 0.$$

6. The systematic sampling of clusters of equal sizes. Let us now suppose that our population consists of clusters of elements, the clusters being of equal size, i.e. containing the same number of elements. To be specific, let the population consist of M clusters, where $M = cm$ and each cluster contains N elements, where $N = kn$. Then, the value of the characteristic being measured for the αth element of the ith cluster may be denoted by $x_{i\alpha}$, and the total of all the elements of the ith cluster may be denoted by x_i. The arithmetic mean of the population is $\bar{x}$, and thus

$$M\bar{x} = \sum_{i=1}^{M} \bar{x}_i.$$

where

$$N_i \bar{x}_i = x_i.$$

a. *Complete enumeration of clusters in sample.* First, suppose that we wish to estimate $\bar{x}$ by $\bar{x}'$, where $\bar{x}'$ is the arithmetic mean of the sample obtained by selecting a systematic sample of m of the M clusters, and enumerating all elements within each cluster in the sample. Then, we may write

$$m\bar{x}' = \sum_{i=1}^{m} \bar{x}'_i, \tag{6.1}$$

where $\bar{x}'_i$ is the mean of the ith cluster selected for the sample. From [1], it follows then that

$$\sigma^2_{\bar{x}'} = \frac{\sigma^2_b}{m} \{1 + (m - 1)\bar{\rho}_c\}$$

where $M\sigma^2_b = \sum_{i=1}^{M} (\bar{x}_i. - \bar{x})^2$, and $\bar{\rho}_c$ is defined as $\bar{\rho}_k$ in [1, p. 6], but with $\bar{x}_i$ in place of x_i. Now from the theory of the random sampling of clusters it follows that

$$\sigma^2_b = \frac{\sigma^2}{N} \{1 + (N - 1)\rho\}$$

where σ^2 is the variance of the population, i. e.

$$MN\sigma^2 = \sum_{i=1}^{M} \sum_{j=1}^{N} (x_{ij} - \bar{x})^2$$

and ρ is the intraclass correlation coefficient of elements within clusters, i. e.

$$\sigma^2 \rho = \sigma^2_b - \sigma^2_w / N - 1,$$

where

$$MN\sigma_w^2 = \sum_{i=1}^{M}\sum_{j=1}^{N} (x_{ij} - \bar{x}_i)^2.$$

Thus

$$\sigma_{\bar{x}'}^2 = \frac{\sigma^2}{mN}\{1 + (N - 1)\rho\}\{1 + (m - 1)\bar{\rho}_c\}. \tag{6.2}$$

Of the three factors in (6.2), σ^2/mN is the variance of a random sample of size mN selected with replacement; $1 + (N - 1)\rho$ is the factor arising from the use of clusters; and $1 + (m - 1)\bar{\rho}_c$ is the factor arising from the fact that the clusters are sampled systematically.

b. *Stratification and subsampling.* When we consider the possibilities of stratification and subsampling, the number of possible designs increases tremendously. For example, it would be simple to calculate the variances of arithmetic means obtained by stratifying the population, selecting sampling units with probability proportionate to size, subsampling systematically, again subsampling systematically and finally subsampling at random. However, such studies may be left to be made in connection with the practical problems in which they are to be used. Rather than attempt to consider many of the possibilities that might arise in practice, we shall here give only the results of the systematic subsampling of a systematic sample. The variances of many other designs may easily be obtained by means of Theorems 1 and 2.

Suppose now that from each of a systematically selected sample of m clusters we subsample, systematically, n elements. Then, let our estimate of $\bar{x}$ be $\bar{x}'$ where, if $x'_{i\alpha}$ is the αth selected element from the ith sample cluster, then

$$\bar{x}' = \left(\frac{1}{mn}\right)\sum_{i=1}^{m}\sum_{\alpha=1}^{n} x'_{i\alpha} = \frac{1}{m}\sum_{i=1}^{m} \bar{x}''_i$$

and

$$\bar{x}''_i = \left(\frac{1}{n}\right)\sum_{\alpha=1}^{m} x'_{i\alpha}.$$

From Theorem 2, it follows at once that

$$\sigma_{\bar{x}'}^2 = \frac{\sigma^2}{mN}\{1 + (N - 1)\rho\}\{1 + (m - 1)\bar{\rho}_c\} + \frac{1}{M}\sum_{i=1}^{M}\frac{\sigma_i^2}{mn}\{1 + (m - 1)\bar{\rho}_i\}, \tag{6.3}$$

where σ_i^2 is the variance within the ith cluster and ρ_i is the average serial correlation within the ith cluster as defined in [1, p. 6]. It is simple to calculate the variance of $\bar{x}'$ also when the sub-sampling is done by considering the m clusters in the sample as one population from which a systematic sample is selected. This is the case that occurs when a sample of blocks is selected and all the households on the sample blocks are listed serially, a systematic sample then being selected from the lists. However, for our present purposes it is the analysis of (6.2) that is important and we now turn to a brief discussion of (6.2).

The most important conclusion to be drawn from (6.2) is that the systematic selection of clusters even when systematic selection is desirable, may not compensate for the increase in variance caused by the use of clusters. Systematic selection will provide the same relative gains but these gains may not be large enough to produce the inequality

$$\{1 + (N - 1)\rho\}\{1 + (m - 1)\bar{\rho}_c\} < \frac{MN - mN}{MN - 1}.$$

A problem that we have not worked through is the following: By regarding the elements of the population as random variables, we obtain conditions on the average correlations among elements of a single cluster as well as on the average correlations among elements of different clusters that enable us to state where the systematic sampling of clusters of equal sizes may be expected to yield a smaller variance than the random or stratified random sampling of clusters or of individual elements. This solution should be straight forward.

c. *Systematic sampling in two dimensions.* Systematic sampling in two dimensions occurs in such practical problems as the selection of a sample of blocks from a city or the selection of a sample of plots from a field.

In selecting blocks from a city, the procedure most often followed effectively reduces the problem to one dimensional form by first numbering the blocks of the city or a part of it, in serpentine fashion beginning, say, in the upper right corner of a map of the city and numbering the blocks in the top row from right to left continuing the numbering of the second row from left to right and so on. Then a systematic sample of these block numbers, and hence, of the blocks themselves is selected. Clearly, this procedure should not be the most efficient if neighboring blocks are highly correlated, since, to cite an unrealistic possibility, the possible samples might turn out to be columns of blocks of the city.

A second two dimensional systematic sampling procedure might be that of selecting a systematic sample of the rows and a systematic sample of the columns, thus obtaining a grid sample. This design too is inefficient when there is a "fertility gradient" along rows or along columns.

The reason for the inefficiency of both of these procedures can be found by examining the formulae for the variances of systematic samples. If the numbering is serpentine, then it becomes illogical to expect that the correlogram is concave upwards and sharp deviations from that pattern may occur. In the grid design, which is a special case of the systematic sampling of clusters with systematic subsampling, we may examine (6.3) and note that the intra-class correlation coefficient ρ may be large enough for $\sigma^2_{\bar{x}'}$ to be large even when $\bar{\rho}_c$ is negative.

Clearly, (6.3) suggests that the possible samples be so defined that ρ is as small as possible. In square fields this might be attained by defining the possible samples to be plots of a Knut Vik square having the same treatment, and similar definitions of possible samples could easily be given for irregular fields. This subject is, however, left for further study.[3]

[3] One of the referees of this paper has drawn the author's attention to an article [6], the data of which, especially Table 3, are in accordance with the opinions expressed above.

REFERENCES

[1] W. G. MADOW AND L. H. MADOW, "On the theory of systematic sampling, I," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 1–24.

[2] L. H. MADOW, "Systematic sampling and its relation to other sampling designs," *Am. Stat. Asso. Jour.*, Vol. 41 (1946), pp. 204–217.

[3] W. G. COCHRAN, "Relative accuracy of systematic and stratified random samples for a certain class of populations," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 164–177.

[4] G. H. HARDY, J. E. LITTLEWOOD AND G. POLYA, *Inequalities*, Cambridge University Press, London and New York, 1934, p. 43.

[5] M. H. HANSEN AND W. N. HURWITZ, "On the theory of sampling from finite populations," *Annals of Math. Stat.*, Vol. 14 (1943), pp. 333–362.

[6] P. G. HOMEYER AND C. A. BLACK, "Sampling replicated field experiments on oats for yield determinations," *Socl. Sci. Soc. Proc.*, Vol. 11 (1946), pp. 341–344.