# SOME INTERESTING RESULTS IN UNEQUAL PROBABILITY SAMPLING FROM A FINITE POPULATION

Grant Capps, U. S. Bureau of the Census

The purpose of this paper is two-fold. First, several results are presented for dealing with the most general of unequal probability sampling schemes. These results are considerably more general than presented in most texts, which generally only deal with the two special cases of unequal probability with and without replacement sampling schemes. The first stage of selection in the Current Population Survey as conducted by the U. S. Bureau of the Census provides a useful application of the general theory. Second, for the common within stratum sample size of n=2, this paper proposes a simple sample selection method that attempts to serve as a compromise between the two frequently opposing survey requirements of a small true variance and an unbiased and fairly stable estimate of that variance. Essentially, this new sampling scheme makes use of both the a-priori information used in the strata formation and a well-known unequal probability without replacement selection method. For the proposed scheme, two estimators of the population total are considered and compared both theoretically and empirically.

## I. GENERALIZED UNEQUAL PROBABILITY SAMPLING FROM A FINITE POPULATION

As is well known, the popular unequal probability with and without replacement sampling schemes are special cases of a much more general sampling scheme. In the following sections, the general theory associated with this general sampling scheme is developed. Please note, it is not claimed that each of the general results about to be presented are necessarily new and original; however, some of these results are at best not very well-known, while others are included for completeness.

A. General Sampling Scheme. Suppose it is required to select a sample for the purpose of estimating some unknown population total. In general, the sampler is free to assign varying probabilities (including zero) to each possible sample configuration. Let there be N population units and suppose we wish to select a sample of size n, not necessarily distinct, units, where n is a fixed constant. The $i^{th}$ population unit has a known variate (or measure of size) $x_i$ and an unknown variate (characteristic of interest) $y_i$ associated with it (i=1=1,2,...,N).

Let $Y = \sum_{1}^{N} y_i$, $X = \sum_{1}^{N} x_i$, and $P_i = x_i/X$ (i=1,2,...,N).

We seek to estimate the unknown population total Y by selecting a sample of size n using some well-defined sampling scheme.

Denote by $t_i$ (i=1,2,...,N) the number of times the $i^{th}$ unit is included in the chosen sample. A technique originally proposed by Cornfield [3], and used by both Cochran [2] and Raj [7] in their excellent sampling tests when handling the special cases of with and without replacement sampling, is to treat the $t_i$ (i=1,...,N) as the random variables rather than the $y_i$ (i=1,2,...,n) where here $y_i$ is the value of the characteristic for the $i^{th}$ unit selected in sample. Raj [7] went slightly further and proposed using the "$t_i$" technique for any general sample design. However, he did not present the relevant results, as done here. The sampling scheme itself will uniquely determine the joint probability distribution of the $t_i$. As will be shown, for the quantities usually estimated, the joint distribution of the $t_i$ is not required. All that is generally required, are the two marginal probabilities, $Pr(t_i)$ and $Pr(t_i,t_j)$ for $i \neq j$, which permit the computation of $E(t_i)$ and $E(t_i t_j)$.

Clearly, since n is fixed, we have

$$n = \sum_{1}^{N} t_i = \sum_{1}^{N} E(t_i) \quad \text{and} \tag{1}$$

$$\mathrm{Cov}(t_i,t_i) = \mathrm{Cov}\left(t_i,\ n - \sum_{j\neq i}^{N} t_j\right) = -\sum_{j\neq i}^{N} \mathrm{Cov}(t_i,t_j) \tag{2}$$

The nature of the sampling scheme employed will determine the difficulty involved in computing $E(t_i)$ and $\mathrm{Cov}(t_i,t_j)$ which are assumed to exist.

B. An Unbiased Estimator for Y. The generalized estimator for Y considered here is:

$$\hat{Y} = \sum_{1}^{N} (y_i) \frac{t_i}{E(t_i)} = \sum_{1}^{n} \frac{y_i}{E(t_i)} \ . \tag{3}$$

The usual assumption that $E(t_i)>0$ (i=1,2,...,N) has been implicily made here. If the variable of interest ($y_i$) and the measures of size ($x_i$) are highly correlated, then we generally desire $E(t_i)$ to be proprotional to $x_i$ (i=1,2,...,N). $\hat{Y}$ is clearly an unbiased estimator for Y, since

$$E(\hat{Y}) = \sum_{1}^{N} (y_i) \frac{E(t_i)}{E(t_i)} = Y.$$

There are many unbiased estimators, some possibly better and others worse, however, this paper addresses only the above estimator, as it is the general extension of the classical with and without replacement (Horvitz-Thompson) [6] estimators. The variance of $\hat{Y}$ and two unbiased variance estimators will now be derived.

C. The Variance of $\hat{Y}$. The sampling variance of $\hat{Y}$ can be expressed in two different, though algebraically, equivalent ways. The straightforward expression for $V(\hat{Y})$ is the quadratic form given by

$$V(\hat{Y}) = \mathrm{Cov}(\hat{Y},\hat{Y}) = \sum_{i}^{N}\sum_{j}^{N} \frac{y_i}{E(t_i)} \frac{y_j}{E(t_j)} \mathrm{Cov}(t_i,t_j) \ . \tag{4}$$

Using (2) $V(\hat{Y})$ can be alternatively expressed as:

$$V(\hat{Y}) = -\sum_{i<j}^{N}\sum^{N} \mathrm{Cov}(t_i,t_j)[\Delta y_{ij}^2], \tag{5}$$

$$\text{where } \Delta y_{ij} = \left[\frac{y_i}{E(t_i)} - \frac{y_j}{E(t_j)}\right].$$

D. Two Unbiased Variance Estimators. Two different variance estimators are suggested by (4) and (5) whenever $E(t_i t_j)>0$ for all distinct pairs $i \neq j$. From (4), it is clear that an unbiased estimator of the sampling variance V(Y)

is given by

$$v_1(\hat{Y}) = \sum_{i\neq j}^{N}\sum^{N} \frac{t_i t_j}{E(t_i t_j)} \frac{y_i}{E(t_i)}\left[\frac{y_j}{E(t_j)}\right] \mathrm{Cov}(t_i,t_j) + \sum_{i}^{N} \frac{t_i}{E(t_i)}\left[\frac{y_i}{E(t_i)}\right]^2 \mathrm{Cov}(t_i,t_i) \quad (6)$$

From (5) it is obvious that another unbiased estimator for $V(\hat{Y})$ is

$$v_2(\hat{Y}) = -\sum_{i\neq j}^{N}\sum^{N} \frac{t_i t_j}{E(t_i t_j)} \mathrm{Cov}(t_i,t_j)\,[\Delta y_{ij}^2] \quad (7)$$

$$= -\sum_{i<j}^{n}\sum^{n} \frac{\mathrm{Cov}(t_i,t_j)}{E(t_i t_j)}\,[\Delta y_{ij}^2] \quad . \quad (8)$$

Only in special cases, such as with simple random sampling, are expressions (6) and (7) equivalent. It should be emphasized that (6) and (7) (or (8)) are unbiased estimators for $V(Y)$ whenever $E(t_i t_j)>0$ for all distinct pairs of population units. If $E(t_i t_j)=0$ for any pair of units, then special assumptions are needed for an unbiased variance estimator to exist.

E. Remark Concerning the Fixed Sample Size n. It should be clear from the above proofs that (3), (4), and (6) are valid for both fixed and random sample sizes, while (5) and (8) do require a fixed sample size, as assumed. Thus, some of the above theory is more general than initially stated.

F. The Stability of the Variance Estimators. When sampling is without replacement, $v(\hat{Y})$ becomes the familiar Horvitz-Thompson [6] variance estimator and $v_2(\hat{Y})$ becomes the well-known Yates-Grundy [9] estimator. In this case, $v_2(\hat{Y})$ is generally the preferred estimator for $V(\hat{Y})$ because it usually is much more stable than $v_1(\hat{Y})$ and assumes negative values less often. Thus, it would seem reasonable to prefer $v_2(\hat{Y})$ over $v_1(\hat{Y})$ in the general scheme. The sampling variance of $v_2(\hat{Y})$ is quite cumbersome; however, when n=2, as is often the case in practice, $v_2(\hat{Y})$ involves only two sample units and the variance of $v_2(\hat{Y})$ is conveniently obtained from (8) as

$$V[v_2(\hat{Y})] = \sum_{i<j}^{N}\sum^{N} \frac{[\mathrm{Cov}(t_i,t_j)]^2}{E(t_i t_j)} (\Delta y_{ij}^4) - [V(\hat{Y})]^2. \quad (9)$$

G. Remark Concerning Multi-Stage Sampling. This section concludes with one final remark. It should be pointed out that the general theory just developed is applicable in two quite different situations. First of all, the general results are obviously valid when dealing with a single stage sample design. In addition, the general theory is also applicable, without modification, for any multi-stage sampling scheme, as long as the sample size at the final stage is fixed. For example, in a multi-stage design, the $y_i$ are the variate values of the final stage units, and n is the fixed number of these final stage units selected for sample. Of course, for computational reasons and because we often wish to know the variances at the various stages, alternative forms for the variance and its estimators showing the several stages of sampling would have to be developed as needed.

## II. A USEFUL APPLICATION OF THE GENERAL THEORY

While it's true that nearly all sample designs in practice are either with or without replacement designs, there does exist at least one ongoing sample survey for which the general theory is quite helpful. The Current Population Survey (CPS) as designed by the U. S. Census Bureau provides us with a useful application of the general theory. The CPS [8] is a stratified multi-stage general population survey of the nation. For simplicity we will focus only on the first stage of selection (for the non-certainty primary units, of course) as if it were the only stage of sampling.

A. The CPS First Stage Sampling Scheme. The sampling scheme used at the first stage of the CPS is certainly an unusual one, and actually resulted from combining two separate existing surveys. We can best describe the sampling scheme for the combined survey as follows. In a typical pair of strata (there are many stratum pairs), choose a sample of size n=3 by initially selecting one stratum at random (i.e., $p=\frac{1}{2}$) and choosing two units with replacement from the chosen stratum using probabilities proportional to some measure of size. Then select one unit with probability proportional to size from the remaining stratum.

B. A Model for Computing the Desired Marginal Probabilities and Expectations. We will now apply the general theory in a typical pair of strata. Let $S_1$ and $S_2$ denote the collection of units in stratum 1 and 2, respectively. For simplicity, assume the first $N_1$ units are in $S_1$ and the last $N_2$ units are in $S_2$ ($N=N_1+N_2$). Let $X_h$ and $Y_h$ (h=1,2) be the stratum totals of the known and unknown variates, respectively. Then,

$$X_h = \sum_{i\in S_h}^{N} x_i \text{ and } Y_h = \sum_{i\in S_h}^{N} y_i \quad (h=1,2).$$

If the $i^{th}$ unit is in stratum h, define the within stratum selection probabilities as

$$p_i' = \frac{x_i}{X_h} \quad (i\in S_h) \text{ for } h=1,2.$$

One simple way to view the selection scheme is to imagine three independent draws from the N units, with one unit selected at each draw. The following three vectors of selection probabilities are used at the various draws:

$$\text{Draw 1: } (p_1', p_2', \ldots, p_{N_1}', 0, 0, \ldots, 0)$$

$$\text{Draw 2: } (0, 0, \ldots, 0, p_{N_1+1}', p_{N_1+2}', \ldots, p_N')$$

$$\text{Draw 3: } \left(\frac{p_1'}{2}, \frac{p_2'}{2}, \ldots, \frac{p_{N_1}'}{2}, \frac{p_{N_1+1}'}{2}, \frac{p_{N_1+2}'}{2}, \ldots, \frac{p_N'}{2}\right)$$

This or any other equivalent model of our sampling scheme allows us to easily compute the following marginal probabilities. For all i we have,

$$\Pr(t_i=1)=\frac{3}{2}p_i' - (p_i')^2, \text{ and } \Pr(t_i=2)=\tfrac{1}{2}(p_i')^2.$$

If the units i and j, $i\neq j$, are in the same stratum, we have

$$\Pr(t_i=1,\ t_j=1) = p_i' p_j' ,$$

while if units i and j are in different strata

$$\Pr(t_i=1,t_j=1) = p_i' p_j'(1-p_i') + p_i' p_j'(1-p_j'), \text{ and}$$

$$\Pr(t_i=1,t_j=2) = \tfrac{1}{2} p_i'(p_j')^2.$$

Using these probabilities the needed expectations are then easily arrived at.

$$E(t_i) = \frac{3}{2} p_i' \quad \text{(all i)},$$

$$E(t_i t_j) = \begin{cases} p_i'\left(\frac{3}{2} + p_i'\right) & i=j \\ p_i' p_j' & \text{i and j in same stratum } (i\neq j) \\ 2p_i'p_j' & \text{i and j in different stratum,} \end{cases}$$

and

$$\mathrm{Cov}(t_i,t_j) = \begin{cases} \frac{p_i'}{2}\left(3-\frac{5}{2} p_i'\right) & i=j \\ -\frac{5}{4} p_i'p_j' & \text{i and j in same stratum } (i\neq j) \\ -\frac{1}{4} p_i'p_j' & \text{i and j in different strata.} \end{cases}$$

These expectations can now be used in conjunction with the general results to derive explicit formulae for $\hat{Y}$, $V(\hat{Y})$, and $v_2(\hat{Y})$.

## III. A NEW COMPROMISE SELECTION METHOD FOR n=2 SAMPLE UNITS PER STRATUM

We now turn to a somewhat unrelated topic concerning efficient survey design. One of the simplest techniques for reducing the variance of an estimator is through effective stratification or universe partitioning. Frequently, due to the large amount of auxiliary information available, stratification may be so effective that it is only necessary to select one sample unit per stratum. However, as is well known, samples of size one generally permit only a positively biased estimate of the variance. Consequently, if there is a pressing need for an unbiased variance estimator, the sampler generally redefines his strata by pairing existing strata and selecting a sample of size two from each new stratum pair. If the sample within each new stratum is chosen in such a way that <u>all</u> pairs of distinct universe units have a positive joint probability of occurrence into the sample, then an unbiased estimate of variance will exist. Unfortunately, there is generally a loss in the actual precision obtained by the latter selection method when compared to the former. Appropriately, this decrease in precision associated with the latter method can sometimes be expressed as a simple function of the bias in the variance estimator used with the former method.

This paper shortly proposes a new selection method for the within stratum sample size n=2. This selection scheme is motivated by the frequent need for an unbiased and stable estimate of the variance of $\hat{Y}$, while at the same time sacrificing as little as possible in the actual sampling variance of $\hat{Y}$, thus resulting in an accurate interval estimate for Y. The proposed method is a simple compromise between a stratified scheme where one unit is selected from each of two strata and, the well-known Durbin [4] selection scheme where two units are selected ignoring stratum boundaries. Two unbiased estimators for Y will be proposed and evaluated, along with their unbiased variance estimators.

A. <u>Stratified Scheme - Scheme 1</u>. The stratified selection scheme will be referred to as Scheme 1. In Scheme 1, the within stratum probabilities, $p_i' = x_i / X_h$ ($i\epsilon S_h$, h=1 or 2), are used in the selection of the two sample units, one from each of the two strata. Denote by $\hat{Y}_s$ the usual unbiased estimator for Y using Scheme 1. Using the earlier results, $\hat{Y}_s$ is given by

$$\hat{Y}_s = \sum_{i}^{N} y_i \frac{t_i}{p_i'} = \sum_{h}^{2} \sum_{i\epsilon S_h}^{N} y_i \frac{t_i}{p_i'} \tag{10}$$

with variance

$$V(\hat{Y}_s) = \sum_{h}^{2} \sum_{\substack{i<j \\ S_h}}^{N} \sum^{N} p_i'p_j'\left(\frac{y_i}{p_i'} - \frac{y_j}{p_j'}\right)^2 . \tag{11}$$

Although Scheme 1 provides us with a precise point estimate of Y, the biased interval estimate it also provides may be unacceptable in certain applications.

1. <u>Special Techniques for Estimating the Variance in Scheme 1</u>. Since $E(t_i t_j)=0$ for all distinct pairs in the same stratum, no unbiased variance estimator exists. A biased, usually positively biased, estimate of variance is obtainable by pairing or collapsing the two strata. Several interesting relationships between the bias in the estimate of variance, the actual variance, and the variance that would have been obtained if a sample of size n=2 had been selected from the N units with replacement will now be developed. Let $\hat{Y}_W$ be the estimator for Y using this with replacement scheme. Applying the general theory yields

$$\hat{Y}_W = \sum_{i}^{N} y_i \frac{t_i}{2p_i} \tag{12}$$

with variance

$$V(\hat{Y}_W) = \sum_{i<j}^{N} \sum^{N} 2p_ip_j\left(\frac{y_i}{2p_i} - \frac{y_j}{2p_j}\right)^2, \tag{13}$$

which upon using (11) becomes

$$V(\hat{Y}_W) = \tfrac{1}{2} V(\hat{Y}_s) + \sum_{\substack{i \\ S_1}}^{N} \sum_{\substack{j \\ S_2}}^{N} 2p_ip_j\left(\frac{y_i}{2p_i} - \frac{y_j}{2p_j}\right)^2 . \tag{14}$$

Suppose the $i^{th}$ unit is selected from stratum 1 and the $j^{th}$ unit is selected from stratum 2. Then $Y_s = \frac{y_i}{p_i'} + \frac{y_j}{p_j'}$. An estimator for $V(\hat{Y}_s)$ that is often used is

$$v(\hat{Y}_s;a_1,a_2) = \left(a_1\frac{y_i}{p_i'} - a_2\frac{y_j}{p_j'}\right)^2 , \tag{15}$$

where $a_1$ and $a_2$ are known constants and are not dependent upon the two units selected for sample. The expectation of $v(\hat{Y}_s;a_1,a_2)$ is

$$Ev(\hat{Y}_s;a_1,a_2) = \sum_{\substack{i \\ S_1}}^{N} \sum_{\substack{j \\ S_2}}^{N} 4p_ip_j\left(a_1\sqrt{\frac{X_1}{X_2}}\frac{y_i}{2p_i} - a_2\sqrt{\frac{X_2}{X_1}}\frac{y_j}{2p_j}\right)^2 \tag{16}$$

and the expectation of its square is

$$E[v(\hat{Y}_s;a_1,a_2)]^2 = \frac{4X_1X_2}{X^2} \sum_{S_i}^{N} \sum_{S_2}^{N} 4p_ip_j$$

$$\times \left( a_1 \sqrt{\frac{X_1}{X_2}} \frac{y_i}{2p_i} - a_2 \sqrt{\frac{X_2}{X_1}} \frac{y_j}{2p_j} \right)^4 . \quad (17)$$

Let us agree to choose $a_1$ and $a_2$ such that $a_1\sqrt{\frac{X_1}{X_2}} = a_2\sqrt{\frac{X_2}{X_1}} = K$. Then using (14), (16) becomes

$$Ev(\hat{Y}_s;a_1,a_2) = 2K^2[V(\hat{Y}_W) - \tfrac{1}{2}V(\hat{Y}_s)], \quad (18)$$

thus showing the bias alluded to earlier, and (17) becomes

$$E[v(\hat{Y}_s;a_1,a_2)]^2 = K^4 \frac{4X_1X_2}{X^2} \sum_{S_1}^{N} \sum_{S_2}^{N} 4p_ip_j$$

$$\times \left( \frac{y_i}{2p_i} - \frac{y_j}{2p_j} \right)^4 . \quad (19)$$

It would be desirable to choose K so that the mean squared error of $v(\hat{Y}_s;a_1,a_2)$, $M[v(\hat{Y}_s;a_1,a_2)] = V[v(\hat{Y}_s;a_1,a_2)] + [Ev(\hat{Y}_s;a_1,a_2)-V(\hat{Y}_s)]^2$, is small. There are three sets of values for $a_1,a_2$, and K sometimes used in practice.

(i) $a_1 = \sqrt{\frac{X_2}{X_1}}$, $a_2 = \sqrt{\frac{X_1}{X_2}}$, and $K^2=1$, in which case

$$Ev(\hat{Y}_s;a_1,a_2) - V(\hat{Y}_s) = 2[V(\hat{Y}_W)-V(\hat{Y}_s)], \quad (20)$$

that is, the bias is equal to twice the (probable) reduction in the actual variance between the two schemes.

(ii) $a_1 = \frac{X}{2X_1}$, $a_2 = \frac{X}{2X_2}$, and $K^2 = \frac{X^2}{4X_1X_2}$. Since $K^2 \geq 1$, this choice of $a_1$ and $a_2$ generally gives a larger bias than does choice 1.

(iii) $a_1 = \frac{2X_2}{X}$, $a_2 = \frac{2X_1}{X}$, and $K^2 = \frac{4X_1X_2}{X^2}$. Since $K^2 \leq 1$ this choice of $a_1$ and $a_2$ generally gives a smaller bias than does choice 1.

In the past, the Bureau has frequently used both the first and third sets of "a" weights $(a_1,a_2)$ as given above.

B. Durbin Scheme - Scheme 2. The Durbin [4] selection scheme will be referred to as Scheme 2. In Scheme 2, the basic selection probabilities, $p_i=x_i/X$, are used in conjunction with the Durbin selection method in selecting two sample units from the two combined strata, completely ignoring the stratum boundaries. The Durbin selection scheme is a simple unequal probability without replacement selection scheme that selects n=2 units per stratum, with inclusion probabilities $\pi_i=2p_i$ and joint inclusion probabilities.

$$\pi_{ij} = \frac{2p_ip_j}{\lambda}\left[\frac{1}{1-2p_j} + \frac{1}{1-2p_i}\right], \quad (i\neq j) \quad (21)$$

where $\lambda = 1 + \sum_{k}^{N} \frac{p_k}{1-2p_k}$. The Durbin method of selection has been shown [1,4] to possess several highly desirable properties. This scheme is used at various stages of selection in several surveys at the Bureau.

Let $\hat{Y}_\pi$ be the usual unbiased estimator for Y obtained from Scheme 2. The previous results show that $\hat{Y}_\pi$ is given by

$$\hat{Y}_\pi = \sum^{2} \frac{y_i}{\pi_i} \quad (22)$$

with variance

$$V(\hat{Y}_\pi) = \sum_{i<j}^{N} \sum^{N} (\pi_i\pi_j-\pi_{ij})\left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2 , \quad (23)$$

with $\pi_i$ and $\pi_{ij}$ as given above.

C. The New Compromise Scheme—Scheme 3. The new compromise selection method, referred to as Scheme 3, will now be given. This new scheme is a simple combination of Schemes 1 and 2, and is motivated by the desire for a selection scheme that possesses most of the optimum properties of these two schemes. Specifically, we desire the (expected) lower variance associated with Scheme 1 and the unbiased and stable variance estimator accompanying Scheme 2.

Let p be any constant satisfying $0\leq p\leq 1$. Then to apply the new compromise Scheme 3, simply choose either Scheme 1 or Scheme 2 with probabilities p and 1-p, respectively, and proceed to select the sample according to the chosen scheme. In an actual survey situation, of course, Scheme 3 would be applied separately in each of many stratum-pairs. Using Scheme 3, two unbiased estimators for Y will be considered along with the variance and unbiased estimate of variance for each. The properties of these estimators and the considerations involved in the choice of p will be the subject of the remainder of this section.

D. Overall Inclusion Probabilities for Scheme 3. Recall that $\pi_i=2p_i=2x_i/X$ and $\pi_{ij}$ are the Scheme 2 (Durbin) inclusion and joint inclusion probabilities, respectively, and that $p_i' = \frac{x_i}{X_h}$ $(i\in S_h)$ are the Scheme 1 inclusion probabilities. Then, if unit i is in stratum h (either h=1 or h=2 for every i), the Scheme 3 inclusion probability is the function of p given by

$$\pi_i(p) = (p_i')p + (\pi_i)(1-p) \quad (i\in S_h, h=1 \text{ or } 2). \quad (24)$$

The Scheme 3 joint inclusion probability for units i and j $(i\neq j)$ is the following function of p:

$$\pi_{ij}(p) = \begin{cases} (p_i'p_j')p + (\pi_{ij})(1-p) & \text{if i,j in different strata} \\ (\pi_{ij})(1-p) & \text{if i,j in same stratum.} \end{cases} \quad (25)$$

Since the Durbin method satisfies $\pi_i\pi_j>\pi_{ij}$ for all $i\neq j$, then if units i and j are in different strata, $\pi_{ij}(p)>\pi_{ij}$. Therefore, it is clear that the effect of Scheme 3, when compared to Scheme 2, is to increase the joint occurrence of units in different strata, while decreasing the joint probability of units in the same stratum.

E. Unconditional Estimator for $Y-\hat{Y}_p$. Under Scheme 3, an unconditional estimator for Y is the usual unbiased estimator, which is p dependent,

and is given by

$$\hat{Y}_p = \sum^{N} (t_i) \frac{y_i}{\pi_i(p)} \qquad (26)$$

with variance

$$V(\hat{Y}_p) = \sum_{i<j}^{N\,N} d_{ij}(p)\,[\Delta y_{ij}(p)]^2 \qquad (27)$$

where $d_{ij}(p) = \pi_i(p)\pi_j(p) - \pi_{ij}(p)$, and

$$\Delta y_{ij}(p) = \left(\frac{y_i}{\pi_i(p)} - \frac{y_j}{\pi_j(p)}\right).$$

Although probably not obvious from (27), $V(\hat{Y}_p)$ is not necessarily monotone (decreasing or increasing) between p=0 and p=1. As we will soon see in the numerical examples, $V(\hat{Y}_p)$ can either be monotone or have peaks and valleys between the two endpoints p=0 and p=1. Thus, one should have sufficient information in order to efficiently specify a value of p when applying Scheme 3.

The unbiased Yates-Grundy estimator for $V(\hat{Y}_p)$ is

$$v(\hat{Y}_p) = \frac{d_{ij}(p)}{\pi_{ij}(p)}\,[\Delta y_{ij}(p)]^2 \,, \quad (p\neq 1) \qquad (28)$$

where the $i^{th}$ and $j^{th}$ units are the selected units.

As can be seen from both the variance estimator $v(\hat{Y}_p)$ in (28) and its variance, $V[v(\hat{Y}_p)]$ obtained obtained from (9), the stability of our variance estimator is dependent upon both p and the effectiveness of the stratification. Although we can't allow p to become too large (near unity), one would expect this scheme can tolerate larger values of p, if desirable, when stratification is effective than if it is not. This is because, although $\pi_{ij}(p)$ is small for units in the same stratum, so also is $[\Delta y_{ij}(p)]^2$ whenever stratification is effective.

In summary, to efficiently apply the unconditional estimator $\hat{Y}_p$ under Scheme 3, one must attempt to find a value of p that jointly produces a small true variance for the estimator of Y, and a stable variance estimator. The criterion used in this paper to quantify the preceding sentence is to find the value of p that minimizes

$$Q_p = V(\hat{Y}_p) + \sqrt{V[v(\hat{Y})_p]} \quad (p\neq 1). \qquad (29)$$

A small value of $Q_p$ should, in some sense, tend to indicate a "good" interval estimate for Y, on the average. Other possible measures of the accuracy of our interval estimate would include differentially weighting each of $Q_p$'s components. The requirements of the survey and the statistician's subjective and objective judgments would ultimately determine these weights.

F. Conditional Estimator for $Y-\hat{Y}_c$. Under Scheme 3, a conditionally (conditioned on the randomly selected scheme) unbiased estimator for Y is given by

$$\hat{Y}_c = \begin{cases} \hat{Y}_s & \text{if Scheme 1 is chosen} \\ \hat{Y}_\pi & \text{if Scheme 2 is chosen} \end{cases} \qquad (30)$$

with variance

$$V_p(\hat{Y}_c) = p\,V(\hat{Y}_s) + (1-p)\,V(\hat{Y}_\pi) \qquad (31)$$

$$= \sum_{i<j}^{N\,N} [\pi_i\pi_j - \alpha_{ij}(p)]\Delta y_{ij}^2 \qquad (32)$$

where $\alpha_{ij}(p)$ is defined by

$$\alpha_{ij}(p) = \begin{cases} \pi_i\pi_j p + (\pi_{ij})(1-p) & \text{if units i,j are in different strata} \\ (\pi_{ij})(1-p) & \text{if units } i\neq j \text{ are in same stratum} \end{cases} \qquad (33)$$

and where $\Delta y_{ij} = \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)$. Note that although $\hat{Y}_c$ does not depend upon p, its sampling distribution does. It is obvious from (31) that, unlike $V(\hat{Y}_p)$, $V_p(\hat{Y}_c)$ is monotone between p=0 and p=1, and further, $V(\hat{Y}_\pi)$ and $V(\hat{Y}_s)$ uniquely determine $V_p(\hat{Y}_c)$. Thus, if stratification is effective, it follows that for all p,

$V_1(\hat{Y}_c) = V(\hat{Y}_s) \leq V_p(\hat{Y}_c) \leq V(\hat{Y}_\pi) = V_0(\hat{Y}_c)$. The unbiased Yates-Grundy type estimator for $V_p(\hat{Y}_c)$ is

$$v_p(\hat{Y}_c) = \frac{\pi_i\pi_j - \alpha_{ij}(p)}{\pi_{ij}(p)}\,[\Delta y_{ij}^2] \quad (p\neq 1) \qquad (34)$$

The comments just made concerning the stability of $v(Y_p)$ hold here for $v_p(Y_c)$ also.

Therefore, when stratification is effective, one should choose p as large as possible, subject to the constraint of a stable variance estimator. The suggested criterion, analogous to the earlier one, is to choose p such that

$$Q_{cp} = V_p(\hat{Y}_c) + \sqrt{V[v_p(\hat{Y}_c)]} \quad (p\neq 1) \qquad (35)$$

is minimized. This optimum value of p should then provide us with an accurate interval estimate for Y.

Scheme 1 is clearly a special case of Scheme 3 and is obtained by simply letting p=1. For this case, both the conditional and the unconditional Scheme 3 estimators become equivalent to the stratified estimator $\hat{Y}_s$, and thus, our criterion for measuing the accuracy of the interval estimates becomes

$$Q_s = V(\hat{Y}_s) + \sqrt{M[v(\hat{Y}_s;a_1,a_2)]} \,, \qquad (36)$$

and is dependent upon the "a" weights chosen.

## IV. TWO NUMERICAL EXAMPLES USING HORVITZ & THOMPSON'S NATURAL POPULATION

In this final section, two numerical examples are considered. For each illustration, the properties of Schemes 1,2, and 3 are explored. As the examples show, the performance of any of the schemes significantly depend upon the population and the quality of the stratification. The first example demonstrates the significant gains obtained by effective stratification, the associated overestimation of the variance, and how Scheme 3 can serve as an effective compromise between Schemes 1 and 2. The second example is included to demonstrate the consequences of ineffective stratification.

A. Horvitz and Thompson's Natural Population. In their 1952 paper, Horvitz and Thompson [6] investigated a universe consisting of N=20 blocks in Ames, Iowa. The data is given in table 1, where the measures $x_i$ are the number of eye-estimated households on the $i^{th}$ block and the $y_i$ are the actual number of households. The data has

been reordered here for clarity. Many authors have subsequently tested their sampling schemes on this population. Table 2 is a summary of results obtained by Horvitz and Thompson [6], Hartley and Rao [5], and Raj [7]. Two numerical examples dealing with Scheme 3 will be given.

Table 1

HORVITZ-THOMPSON NATURAL POPULATION

| i | $y_i$ | $x_i$ | $y_i/x_i$ |
|---|---|---|---|
| 1 | 19 | 18 | 1.06 |
| 2 | 9 | 9 | 1.00 |
| 3 | 21 | 24 | .88 |
| 4 | 22 | 25 | .88 |
| 5 | 15 | 14 | 1.07 |
| 6 | 18 | 18 | 1.00 |
| 7 | 37 | 40 | .93 |
| 8 | 12 | 12 | 1.00 |
| 9 | 27 | 27 | 1.00 |
| 10 | 25 | 26 | .96 |
| 11 | 19 | 19 | 1.00 |
| 12 | 12 | 12 | 1.00 |
| 13 | 17 | 14 | 1.21 |
| 14 | 14 | 12 | 1.17 |
| 15 | 27 | 23 | 1.17 |
| 16 | 20 | 17 | 1.18 |
| 17 | 25 | 21 | 1.19 |
| 18 | 35 | 24 | 1.46 |
| 19 | 47 | 30 | 1.57 |
| 20 | 13 | 9 | 1.44 |
| | Y=434 | X=394 | |

Table 2

SUMMARY OF PREVIOUS RESULTS

| Sampling Scheme | Variance of the Estimator | $\left(\text{Variance of Variance Estimator}\right)^{\frac{1}{2}}$ |
|---|---|---|
| 1. Simple Random Sampling | 17,122 | 26,539 |
| 2. Stratified Random; one element from each of two strata with equal probability[1] | 7,873 | NA |
| 3. Equal Probability Systematic Sampling | 10,224 | NA |
| 4. pps With Replacement | 3,247 | 4,611 |
| 5. First Horvitz-Thompson Scheme ($\pi$ps)[2] | 3,095 | NA |
| 6. Second Horvitz-Thompson Scheme ($\pi$ps)[3] | 3,075 | NA |
| 7. Systematic $\pi$ps | 3,014 | 3,983 |

NA = Not Available

[1]Stratum 1 consists of the 10 blocks with the largest measures of size ($x_i \geq 19$), with the smallest ($x_i \leq 18$) 10 blocks in stratum 2.

[2]First sample unit selected with pps, second unit from remainder with equal probabilities. Original measures altered so as to obtain an approximate $\pi$ps (i.e., $\pi_i = 2x_{i/X}$) scheme.

[3]First sample unit selected with pps, second unit with pps of the remaining units. Original measures altered so as to obtain an approximate $\pi$ps scheme.

B. Example 1-First Stratification. In the first example, units 1 through 12 comprise stratum 1 and units 13 through 20 comprise stratum 2. From table 1 this would appear to be an effective stratum formation. We have $N_1=12$, $N_2=8$, $X_1=244$, $X_2=150$, $Y_1=236$, and $Y_2=198$.

The results appear in tables 3a and 3b and are encouraging. In this example, the unconditional estimator yields excellent point and interval estimates for any p satisfying $.25 \leq p \leq .65$. In this case the precision obtained compares well with that of Scheme 1. We also see that $V(Y_p)$ behaves quite smoothly in this first stratification. As shown in table 3b, the bias in each of the Scheme 1 variance estimators is probably intolerable to most. In fact, this bias is so sizeable that for nearly each value of p not too near unity, both the unconditional and the conditional estimators are superior to the Scheme 1 estimator when applying the "Q" criterion. Finally, for each p, the unconditional estimator is always superior to the conditional estimator.

C. Example 2-Second Stratification. The effectiveness of the stratification is an important issue, as this second and final example demonstrates. Horvitz and Thompson suggest stratifying according to the measures of size ($x_i$), with the 10 largest eye-estimated blocks in stratum 1 and the 10 smallest in stratum 2. When sampling is with equal probabilities, this method of stratification has already been considered, as shown in table 2 (No. 2). In this case, there was a significant improvement compared to unrestricted simple random sampling (table 2, No. 1). As we will see, such is not the case when comparing stratified unequal probability sampling (using the strata definition just given) with (unrestricted) pps with replacement sampling (table 2, No. 4). Thus, in this example, all units with $x_i \geq 19$ (i=3,4,7,9,10,11,15,17,18 and 19) are defined as stratum 1, and all units with $x_i \leq 18$ (i=1,2,5,6,8,12,13,14,16, and 20) comprise stratum 2. The summary totals are now $N_1=N_2=10$, $X_1=259$, $X_2=135$, $Y_1=285$, and $Y_2=149$.

Inspection of the $y_{i/x_i}$ column in table 1 tend to indicate this second stratification is not very effective. The stratified scheme yields considerably less precision ($V(Y_s)=4025$) than either the Durbin scheme or pps with replacement sampling. Because the stratification was so poor, the precision of both the conditional and the unconditional estimators get steadily worse as $0 \to p \to 1$, although the unconditional estimator begins to dip back down at about $p \doteq .75$. The precision of both variance estimators become steadily worse as $0 \to p \to 1$ because the small joint inclusion probabilities are not being associated with small $[\Delta y_{ij}(P)]^2$, again due to poor stratifying. In addition, each of the three Scheme 1 variance estimators seriously underestimates (2438, 2739, and 2224) the actual variance, whereas when stratification is effective they are generally each overestimates of variance. Therefore, as this example indicates, the quality (or lack of quality) of the stratification is a crucial issue, and, in particular, stratifying only on the basis of size is certainly questionable. The tables showing the analysis for this second example can be obtained upon writing the author.

TABLE 3a

| Horvitz-Thompson Population - First Stratification Results | | | | | | |
|---|---|---|---|---|---|---|
| | Unconditional Estimator | | | Conditional Estimator | | |
| Scheme 3 | $V(\hat{Y}_p)$ | $\sqrt{V[v(\hat{Y}_p)]}$ | $Q_p$ | $V_p(\hat{Y}_c)$ | $\sqrt{V[v_p(\hat{Y}_c)]}$ | $Q_{cp}$ |
| p=0 (Durbin, Scheme 2) | 3011 | 3990 | 7001 | 3011 | 3990 | 7001 |
| p=.10 | 2220 | 2809 | 5029 | 2791 | 3382 | 6173 |
| p=.25 | 1438 | 2042 | 3480 | 2463 | 2717 | 5180 |
| p=.50 | 896 | 2544 | 3440 | 1915 | 2516 | 4431 |
| p=.65 | 842 | 3356 | 4198 | 1586 | 3129 | 4715 |
| p=.75 | 853 | 4202 | 5055 | 1367 | 3963 | 5330 |
| p=.90 | 863 | 7147 | 8010 | 1038 | 6992 | 8030 |
| p=1 (Stratified, Scheme 1) | 819 | -- | -- | 819 | -- | -- |

TABLE 3b

| Horvitz-Thompson Population - First Stratification Results | | | | | |
|---|---|---|---|---|---|
| | $a_1$ | $a_2$ | $Ev(\hat{Y}_s;a_1,a_2)$ | $\sqrt{M[v(\hat{Y}_s;a_1,a_2)]}$ | $Q_s$ |
| 1. | $\sqrt{X_{2/X_1}}$ | $\sqrt{X_{1/X_2}}$ | 5674 | 6984 | 7803 |
| 2. | $X/_{2X_1}$ | $X/_{2X_2}$ | 6017 | 7440 | 8259 |
| 3. | $2X_{2/X}$ | $2X_{1/X}$ | 5351 | 6554 | 7373 |

REFERENCES


1. C. Asok and B. V. Sukhatme. On Sampford's Procedure of Unequal Probability Sampling Without Replacement. Journal of the American Statistical Association, (1976), Vol. 71, pp. 912-918.
2. W. G. Cochran. Sampling Techniques. 2nd ed. New York: Wiley and Sons, 1963.
3. J. Cornfield. On Samples from Finite Populations. Journal of the American Statistical Association, (1944), Vol. 39, pp. 236-239.
4. J. Durbin. Design of Multi-stage Surveys for the Estimation of Sampling Errors. Applied Statistics, (1967), Vol. 16, pp. 152-164.
5. H. O. Hartley and J. N. K. Rao. Sampling With Unequal Probabilities and Without Replacement. Annals of Mathematical Statistics, (1962), Vol. 33, pp. 350-374.
6. D. G. Horvitz and D. J. Thompson. A Generalization of Sampling Without Replacement From a Finite Universe. Journal of the American Statistical Association, (1952) Vol. 47, pp. 663-685.
7. D. Raj. Sampling Theory. 1st ed. New York: McGraw-Hill, 1968.
8. M. Thompson and G. Shapiro. The Current Population Survey: An Overview. Annals of Economic and Social Measurement, (1973), Vol. 2, No. 2.
9. F. Yates and P. M. Grundy. Selection Without Replacement From Within Strata With Probability Proportional to Size. Journal of the Royal Statistical Society, (1953), Series B, Vol. 15, pp. 253-261.


ACKNOWLEDGEMENTS


The author would like to thank Gary Sparks for his excellent computer programming, Kirk Wolter for some helpful comments, and Edith Oechsler for her conscientious typing, all of the Census Bureau. Thanks also to Dr. Harry Rosenblatt of the American University, Washington, D. C., for reviewing a larger version of this paper.