ORNL/TM-10398

ornl

OAK RIDGE NATIONAL LABORATORY

MARTIN MARIETTA

A Simple Method for Probability Proportional to Size (πps) Sampling Without Replacement Based on Ranks

Tommy Wright

OPERATED BY
MARTIN MARIETTA ENERGY SYSTEMS, INC.
FOR THE UNITED STATES
DEPARTMENT OF ENERGY

Printed in the United States of America. Available from
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Road, Springfield, Virginia 22161
NTIS price codes—Printed Copy: A03; Microfiche A01

ORNL/TM-10398

Engineering Physics and Mathematics Division

Mathematical Sciences Section

A SIMPLE METHOD FOR PROBABILITY PROPORTIONAL TO SIZE (πps) SAMPLING WITHOUT REPLACEMENT BASED ON RANKS

Tommy Wright

Date Published: June 1987

Research was supported by the Applied Mathematical Sciences Research Program of the Office of Energy Research, U.S. Department of Energy.

Prepared by the
Oak Ridge National Laboratory
Oak Ridge, Tennessee 37831
operated by
Martin Marietta Energy Systems, Inc.
for the
U.S. DEPARTMENT OF ENERGY
under Contract No. DE-AC05-84OR21400

CONTENTS

A SIMPLE METHOD FOR PROBABILITY PROPORTIONAL TO SIZE (πps) SAMPLING WITHOUT REPLACEMENT BASED ON RANKS

Tommy Wright

Mathematical Sciences Section
Engineering Physics and Mathematics Division
Oak Ridge National Laboratory
P.O. Box Y, Bldg. 9207A, MS 2
Oak Ridge, Tennessee 37831 U.S.A

ABSTRACT

There can be gains in estimation efficiency over equal probability and unequal probability with replacement sampling methods when one makes use of auxiliary information for probability proportional to size without replacement ($\pi pswor$) sampling methods. Such methods are often complex to execute, particularly when the sample size n exceeds 2. Also, appropriate auxiliary information may not be available. When the population units can be ranked according to the variable of interest, we demonstrate that there can be gains in estimation efficiency with a very simple $\pi pswor$ sampling strategy based on ranks for all sample sizes n where the population size $N=2kn$ (k a positive integer).

KEY WORDS: Horvitz-Thompson Estimator; Jackknife; $\pi pswor$ Sampling; Ranks; Variance Estimation.

1. INTRODUCTION

When sampling from a finite population to estimate some unknown parameter, it is often desirable to have unequal probabilities of inclusion for membership in the sample. Let $U_1, U_2, \cdots, U_N$ be the N units of a finite population and assume that a vector (y_i, x_i) of nonnegative components is associated with the i^{th} unit. If the x's are known and the y's are unknown for $i = 1, 2, \cdots, N$, assume that a sample of fixed size n $(< N)$ is to be selected to estimate the unknown population total $Y = \sum_{i=1}^{N} y_i$. Let $\delta_i = 1$ if the i^{th} unit is included in the sample and 0 otherwise. Also let π_i be the probability that the i^{th} unit is included in the sample. Assume that it is desired that $\pi_i = cp_i$ for some $c > 0$ where $p_i = x_i / \sum_{i=1}^{N} x_i$ is a normed measure of size for the i^{th} unit. Then because the expected sample size is restricted by

$$n = E\left(\sum_{i=1}^{N} \delta_i\right) = \sum_{i=1}^{N} \pi_i = \sum_{i=1}^{N} cp_i \ ,$$

we must have $c = n$. Thus if $\pi_i = np_i$ for $i = 1, \cdots, N$, the sampling plan is called *probability proportional to size* (πps) *sampling*.

Hansen and Hurwitz (1943) demonstrated that more efficient estimators of Y can often be obtained using πps sampling, which gives unequal probabilities of sample inclusion, than can be obtained from equal probability sampling. The method of selection that they considered was easy to execute and led to unbiased estimators and unbiased estimators of variance. However, their method was *with* replacement, i.e., a unit could be included in the sample more than once.

A number of *without* replacement πps sampling plans have been proposed which limit each population unit to at most one appearance in a sample, but they are not without problems and limitations. Brewer and Hanif (1983) give a long list of such procedures and compare them with considerable discussion devoted to their strengths and weaknesses. In addition to each unit being included in the sample at most one time, πps *sampling without replacement* ($\pi pswor$) can lead to more efficient estimators than πps *sampling with replacement* ($\pi pswr$). Also, the availability of the Horvitz-Thompson (1952) estimator simplifies the estimation for many of the $\pi pswor$ sampling plans. The known $\pi pswor$ sampling plans all seem to suffer from at least one of the following weaknesses.

(i) The method is extremely complicated to execute if $n > 2$, and it is difficult to compute probabilities of inclusion, especially the joint probabilities of inclusion as in Brewer (1963).

(ii) If a simple method is proposed for $n > 2$, it will often impose very restrictive requirements on the p_i's for general applications as do the methods of Midzuno (Brewer and Hanif, 1983, p. 25), Chaudhuri (1975), and Sengupta (1987).

(iii) Other simple plans of unequal probability sampling without replacement such as those of Rao, Hartley, and Cochran (1962) and Murthy (1957) yield π_i's which are not equal to np_i. Hence these methods are not truly *πpswor* sampling plans.

(iv) Other simple plans like that of Madow (1949) make unbiased estimation of the sampling variance impossible because of systematic selection, not to mention the other biases that can come with systematic selection due to the ordering of the population.

Aside from improvements in efficiency over equal probability sampling, perhaps the chief advantage of *πpswr* sampling is its simplicity in terms of sampling and unbiased estimation, and its major disadvantage is that any unit can appear in the sample more than once. Quite the opposite is true of *πpswor* sampling methods as noted in the previous paragraph. The challenge has been to present an efficient *πpswor* sampling method that is simple in terms of sampling and estimation, particularly for $n > 2$. The feature that contributes to the simplicity of *πpswr* sampling is that one makes n independent selections from the *same* population. By "same" we mean that the population conditions for the i^{th} selection for the sample are no different from the population conditions for the $(i+1)^{st}$ selection for the sample $(i=1,2,\cdots,n)$ in a successive selection process. Is it possible to make n independent selections from the same population without replacement? One way that we can come close to achieving this is to stratify the population into n strata which are each smaller-sized "copies" of the population and select one unit independently from each stratum. That is, we desire that each stratum have similar characteristics as the population—for example, each stratum mean might be required to be the same as (or close to) the population mean. Hence rather than making n independent selections from the same population, we consider the independent selection of one unit from each of n strata where each stratum is a close miniature copy of the population.

Ideally x should be chosen to be a variable that is directly proportional to the unknown variable of interest y. While a suitable variable x is frequently available or can

be cheaply obtained, occasions exist where one may not know a suitable x. Recalling that x is often thought of as a "measure of size" for each unit relative to y, one may consider taking the ranks (of the units based on the unknown y's and on prior information) as the variable x. Many examples exist where one may not know exact y values, but may know how to rank units reasonably well relative to the unknown y values before sample selection. Ranks can be the basis for the construction of the n heterogeneous groups where each is a miniature copy of the overall population. In what follows, we assume that such ranking is possible.

Before proceeding, it is instructive to cite an example where the ranks of the unknown y-values would likely be known for the entire population. There are 160 distributors of the electricity generated by the Tennessee Valley Authority (TVA) for more than 2.5 million residential customers over an eight state region. As part of its Load Research Program, TVA wants to estimate its total demand from these residential customers (as well as other categories of customers) for given time periods. For administrative convenience, multi-stage cluster sampling plans that treat each distributor as the primary sampling unit seem most appropriate for attaining timely estimates of demand. The first stage of such a sampling plan would generally call for the selection of a sample of distributors using $\pi pswor$. The relative total customer demand by distributor is fairly constant from month to month and is known. Hence in this case, if y_i is the total demand of the residential customers of the i^{th} distributor for a given time period, its rank in the list for $N = 160$ distributors would be known with a reasonable confidence assuming fairly uniform demand per distributor over a month.

In this paper, a new and extremely simple method of sample selection is proposed for $\pi pswor$ sampling based on ranks, i.e., we take

$$x_i = \text{the rank of } U_i \tag{1.1}$$

as our measure of size which we assume to be highly correlated with the unknown y_i's. In addition to the method of selection, we exhibit an unbiased estimator (Horvitz-Thompson estimator) of y whose sampling error is no greater than that under $\pi pswr$ and show that it does not suffer from the first three weaknesses mentioned in the previous paragraph. Unbiased estimation of sampling error (variance) is impossible because some joint inclusion probabilities are zero. Therefore, to estimate the sampling variance, we consider a jackknife estimator and obtain an expression for its bias. For some selected populations, numerical comparisons are made between the method of this paper and the simple methods of Rao, Hartley, and Cochran (1962), and Samiuddin and Asad (1981) which are similar in spirit to the method of this paper.

2. THE SAMPLING SCHEME AND ESTIMATION

2.1 GROUP CONSTRUCTION

Without loss of generality, assume that the rank of U_i is $R_i = i$, where the ordering is from smallest to largest with respect to the unknown y_i's. The normed measure of size associated with U_i is therefore

$$p_i = \frac{R_i}{\sum_{j=1}^{N} R_j} = \frac{2i}{N(N+1)} \tag{2.1}$$

Assume further that $N = 2kn$ where k is a positive integer and n is the desired sample size. Partition the N units into n groups each of size $2k$ units as follows:

1. Assign to Group 1 those units with the k smallest ranks and the k largest ranks; i.e., $U_1, U_2, \cdots, U_k, U_{N-k+1}, \cdots, U_{N-2}, U_{N-1}, U_N$.

2. Assign to Group 2 those units with the next k smallest ranks and the next k largest ranks; i.e., $U_{k-1}, \cdots, U_{2k}, U_{N-2k+1}, \cdots, U_{N-k-1}, U_{N-k}$.

   ⋮

n. Assign to group n the middle $2k$ units, i.e.,

$$U_{(n-1)k+1}, \cdots, U_{nk}, U_{nk+1}, \cdots, U_{(n+1)k} .$$

It is easy to see that the sum of the ranks in each group is $N(N+1)/2n$. This stratification will tend to make the strata (or groups) all have means that are close to that of the overall population. This is especially true when there is symmetry in the distribution of the y-values.

2.2 THE SELECTION SCHEME

Let i_j range over the ranks assigned to the j^{th} group and select the $i_j{}^{th}$ unit from group j with probability np_{i_j}, where p_{i_j} is given according to (2.1) for $j = 1, 2, \cdots, n$. The selection of the one unit from a given group is independent of the selection of another unit from another group. Note that the sum of the probabilities in the j^{th} group is

$$\sum_{i_j} np_{i_j} = \frac{2n}{N(N+1)} \sum_{i_j} i_j = 1$$

and that the probability of inclusion in the sample for the i^{th} unit of the population is

$$\pi_i = np_i \quad . \tag{2.2}$$

Hence the sample selection procedure is $\pi pswor$ sampling. So that np_{i_j} will be a probability, it is also necessary that

$$1 > n(\max p_{i_j}) = n\left(\frac{2N}{N(N+1)}\right) = \frac{2n}{N+1} \quad .$$

Thus n must be less than half the population size plus one, a very mild requirement.

If π_{ij} is the joint probability that U_i and U_j are included in the sample, then

$$\pi_{ij} = \begin{cases} 0 & \text{if } U_i \text{ and } U_j \text{ are in the same group} \\ \pi_i \pi_j & \text{if } U_i \text{ and } U_j \text{ are in different groups} \end{cases} \quad . \tag{2.3}$$

2.3 ESTIMATION

To estimate $Y = \sum_{i=1}^{N} y_i$, we use the unbiased Horvitz-Thompson estimator

$$\hat{Y}_{HTWOR} = \sum_{j=1}^{n} \frac{y_j}{\pi_j} \tag{2.4}$$

where y_j is the y-value of the one unit selected from the j^{th} group and π_j is its probability of inclusion.

Lemma 1. Let μ_{jy} be the true mean of the $2k$ units assigned to the j^{th} group. Then

$$Var(\hat{Y}_{HTWOR}) = \sum_{i=1}^{N} \frac{y_i^2}{\pi_i} - \left(\frac{N}{n}\right)^2 \sum_{j=1}^{n} \mu_{jy}^2 \quad . \tag{2.5}$$

Proof. Let δ_i be as defined in the Introduction for $i = 1, \cdots, N$. Clearly

$$E(\delta_i) = \pi_i \quad ,$$
$$Var(\delta_i) = \pi_i(1 - \pi_i)$$

and for $i \neq j$,

$$Cov(\delta_i, \delta_j) = \begin{cases} -\pi_i \pi_j & \text{if } i \text{ and } j \text{ are in the same group} \\ 0 & \text{if } i \text{ and } j \text{ are in different groups} \end{cases} \quad . \tag{2.6}$$

Then $\hat{Y}_{HTWOR}$ can be written as

$$\hat{Y}_{HTWOR} = \sum_{i=1}^{N} \frac{\delta_i y_i}{\pi_i} \quad \text{and}$$

$$Var(\hat{Y}_{HTWOR}) = Var\left[\sum_{i=1}^{N} \frac{\delta_i y_i}{\pi_i}\right]$$

$$= Var\left[\sum_{\substack{i_1 \\ Group\ 1}} \frac{\delta_{i_1} y_{i_1}}{\pi_{i_1}} + \sum_{\substack{i_2 \\ Group\ 2}} \frac{\delta_{i_2} y_{i_2}}{\pi_{i_2}} + \cdots + \sum_{\substack{i_n \\ Group\ n}} \frac{\delta_{i_n} y_{i_n}}{\pi_{i_n}}\right]$$

which becomes by independent sampling among the groups

$$= Var\left[\sum_{\substack{i_1 \\ Group\ 1}} \frac{\delta_{i_1} y_{i_1}}{\pi_{i_1}}\right] + Var\left[\sum_{\substack{i_2 \\ Group\ 2}} \frac{\delta_{i_2} y_{i_2}}{\pi_{i_2}}\right] + \cdots + Var\left[\sum_{\substack{i_n \\ Group\ n}} \frac{\delta_{i_n} y_{i_n}}{\pi_{i_n}}\right]$$

$$= \left[\sum_{\substack{i_1 \\ Group\ 1}} \left(\frac{y_{i_1}}{\pi_{i_1}}\right)^2 Var(\delta_{i_1}) + \sum\sum_{\substack{i_1 \neq j_1 \\ Group\ 1}} \left(\frac{y_{i_1}}{\pi_{i_1}}\right)\left(\frac{y_{j_1}}{\pi_{j_1}}\right) Cov(\delta_{i_1}, \delta_{j_1})\right]$$

$$+ \cdots + \left[\sum_{\substack{i_n \\ Group\ n}} \left(\frac{y_{i_n}}{\pi_{i_n}}\right)^2 Var(\delta_{i_n}) + \sum\sum_{\substack{i_n \neq j_n \\ Group\ n}} \left(\frac{y_{i_n}}{\pi_{i_n}}\right)\left(\frac{y_{j_n}}{\pi_{j_n}}\right) Cov(\delta_{i_n}, \delta_{j_n})\right]$$

$$= \sum_{i=1}^{N} \frac{(1-\pi_i)}{\pi_i} y_i^2 - \sum\sum_{\substack{i_1 \neq j_1 \\ Group\ 1}} y_{i_1} y_{j_1} - \sum\sum_{\substack{i_2 \neq j_2 \\ Group\ 2}} y_{i_2} y_{j_2} - \cdots - \sum\sum_{\substack{i_n \neq j_n \\ Group\ n}} y_{i_n} y_{j_n}$$

$$= \sum_{i=1}^{N} \frac{y_i^2}{\pi_i} - \sum_{i=1}^{N} y_i^2 - \sum\sum_{\substack{i_1 \neq j_1 \\ Group\ 1}} y_{i_1} y_{j_1} - \sum\sum_{\substack{i_2 \neq j_2 \\ Group\ 2}} y_{i_2} y_{j_2} - \cdots - \sum\sum_{\substack{i_n \neq j_n \\ Group\ n}} y_{i_n} y_{j_n}$$

$$= \sum_{i=1}^{N} \frac{y_i^2}{\pi_i} - \left[\sum_{\substack{i_1 \\ Group\ 1}} y_{i_1}\right]^2 - \left[\sum_{\substack{i_2 \\ Group\ 2}} y_{i_2}\right]^2 - \cdots - \left[\sum_{\substack{i_n \\ Group\ n}} y_{i_n}\right]^2$$

$$= \sum_{i=1}^{N} \frac{y_i^2}{\pi_i} - \frac{N^2}{n^2} \sum_{j=1}^{n} \mu_{j_y}^2 \text{ , which proves (2.5).}$$

Let $\hat{Y}_{HTWR}$ be the usual unbiased estimator of Y under $\pi pswr$ and using R_i for x_i. From Theorem 9A.1 of Cochran (1977), it is known that

$$\hat{Y}_{HTWR} = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{p_i}$$

and

$$Var(\hat{Y}_{HTWR}) = \sum_{i=1}^{N} \frac{y_i^2}{\pi_i} - \frac{\left[\sum_{i=1}^{N} y_i\right]^2}{n} \tag{2.7}$$

where $\pi_i = np_i$. The following lemma demonstrates that in terms of sampling variance, the method of this paper using the ranks is always as good or better than $\pi pswr$ using the ranks.

Lemma 2. $Var(\hat{Y}_{HTWOR}) \leq Var(\hat{Y}_{HTWR})$.

Proof. From (2.5) and (2.7),

$$Var(\hat{Y}_{HTWR}) - Var(\hat{Y}_{HTWOR}) = \frac{N^2}{n^2} \sum_{j=1}^{n} \mu_{jy}^2 - \frac{\left[\sum_{i-1}^{N} y_i\right]^2}{n}$$

$$= \frac{N^2}{n^2}\left[\sum_{j=1}^{n} \mu_{jy}^2 - n\,\mu_y^2\right]$$

$$= \frac{N^2}{n^2}\left[\sum_{j=1}^{n} (\mu_{jy} - \mu_y)^2\right]$$

$$\text{because } \mu_y = \frac{Y}{N} = \frac{\sum_{j=1}^{n} \mu_{jy}}{n}$$

$$\geq 0 .$$

This completes the proof of Lemma 2.

3. ESTIMATION OF $Var(\hat{Y}_{HTWOR})$

Unbiased estimators of the sampling variance of the usual Horvitz-Thompson estimator are well known. One estimator by Horvitz and Thompson (1952) and another by Sen (1953) and Yates and Grundy (1953) are referenced widely. These estimators can be negative, and they require that all π_{ij} be nonzero for $i \neq j$ and $i, j = 1, \cdots, N$. Because π_{ij} can be zero for the method proposed in this paper, these two estimators cannot be considered.

Because of the structure of the sampling method and the fact that the jackknife estimator of variance under *πpswor* reduces to the usual unbiased estimator $\hat{Var}(\hat{Y}_{HTWR})$, we consider estimation of $Var(\hat{Y}_{HTWOR})$ by the jackknife method as discussed in section 4.3.4 of Wolter (1985). For $j = 1, 2, \cdots, n$, let

$$\hat{Y}_{(j)} = \sum_{\substack{i=1 \\ i \neq j}}^{n} \frac{y_i}{(n-1)p_i} = \sum_{\substack{i=1 \\ i \neq j}}^{n} \frac{y_i}{\left[\frac{n-1}{n}\right]\pi_i} \tag{3.1}$$

which is the Horvitz-Thompson estimator based on the sample after removing the unit

selected from the j^{th} group using $\left(\frac{n-1}{n}\right)\pi_i$ in place of π_i. For $j=1, \cdots, n$, let

$$\hat{Y}_j = n\hat{Y}_{HTWOR} - (n-1)\hat{Y}_{(j)} . \tag{3.2}$$

An estimator (Quenouille's) of Y is

$$\hat{Y}_J = \frac{\sum_{j=1}^{n} \hat{Y}_j}{n} . \tag{3.3}$$

Note that

$$\hat{Y}_J = \frac{\sum_{j=1}^{n} \hat{Y}_j}{n} = \frac{\sum_{j=1}^{n} (n\hat{Y}_{HTWOR} - (n-1)\hat{Y}_{(j)})}{n}$$

$$= \frac{n^2\hat{Y}_{HTWOR} - n \sum_{j=1}^{n} \sum_{\substack{i=1 \\ i \neq j}}^{n} \frac{y_i}{\pi_i}}{n}$$

$$= \frac{n^2\hat{Y}_{HTWOR} - n(n-1) \sum_{j=1}^{n} \frac{y_j}{\pi_j}}{n}$$

$$= \hat{Y}_{HTWOR} .$$

Thus the estimator $\hat{Y}_J$ is the same as $\hat{Y}_{HTWOR}$. To estimate $Var(\hat{Y}_J)$, which then is the same as $Var(\hat{Y}_{HTWOR})$, the jackknife estimator is

$$\hat{Var}_J(\hat{Y}_J) = \frac{1}{n(n-1)} \sum_{j=1}^{n} (\hat{Y}_j - \hat{Y}_J)^2 \tag{3.4}$$

Therefore our estimator of $Var(\hat{Y}_{HTWOR})$ is

$$\hat{Var}(\hat{Y}_{HTWOR}) = \hat{Var}_J(\hat{Y}_J) . \tag{3.5}$$

Note that $\hat{Y}_j$ reduces to y_j/p_j. Hence

$$\hat{Var}(\hat{Y}_{HTWOR}) = \frac{1}{n(n-1)} \sum_{j=1}^{n} (\hat{Y}_j - \hat{Y}_J)^2$$

$$= \frac{1}{n(n-1)} \sum_{j=1}^{n} \left(\frac{y_j}{p_j} - \hat{Y}_{HTWOR}\right)^2 \tag{3.6}$$

$$= \hat{Var}(\hat{Y}_{HTWR})$$

That is, $\hat{Var}(\hat{Y}_{HTWOR})$, which is the jackknife estimator, reduces to $\hat{Var}(\hat{Y}_{HTWR})$ which is the usual unbiased estimator of $Var(\hat{Y}_{HTWR})$ given in (2.7). Hence the jackknife estimator tells us to pretend that the sampling is with replacement when estimating $Var(\hat{Y}_{HTWOR})$. The following lemma shows that $\hat{Var}(\hat{Y}_{HTWR})$ is a non-negatively biased estimator of $Var(\hat{Y}_{HTWOR})$.

Lemma 3. If $\hat{Var}(\hat{Y}_{HTWR})$ is used to estimate $Var(\hat{Y}_{HTWOR})$, then

$$Bias\ (\hat{Var}(\hat{Y}_{HTWR})) = \frac{N^2}{(n-1)n} \sum_{j=1}^{n} (\mu_{jy} - \mu_y)^2 \tag{3.7}$$

Proof. From Theorem 2.4.6 of Wolter (1985), we observe that

$$Bias\,(\hat{Var}(\hat{Y}_{HTWR})) = \frac{n}{n-1}[Var(\hat{Y}_{HTWR}) - Var(\hat{Y}_{HTWOR})]$$

From the proof of Lemma 2, this reduces to

$$= \frac{N^2}{(n-1)n} \sum_{j=1}^{n} (\mu_{jy} - \mu_y)^2 \ .$$

The $Bias\,(\hat{Var}(\hat{Y}_{HTWR}))$ tends to be small when the μ_{jy}'s are all essentially the same. Clearly if the ranks are directly proportional to the y-values, then $Bias\,(\hat{Var}(\hat{Y}_{HTWR})) = 0$ and $\hat{Var}(\hat{Y}_{HTWR})$ would be an unbiased estimator of $Var(\hat{Y}_{HTWOR})$ for the method proposed in this paper.

4. NUMERICAL COMPARISONS

Sampling with probability proportional to size is most easy and simple when $n = 1$. Clearly in this case there is no difference between *πpswr* and *πpswor* sampling. Recall that a motivating reason for the proposed method of this paper was to stratify the population into n groups so that one independent selection could be made from each group. This simplicity is a chief advantage in terms of execution and is a feature that is shared by the methods of Rao, Hartley, and Cochran (1962) and Samiuddin and Asad (1981). In this section we give numerical comparisons of the method of this paper with four other methods, including the two just mentioned, using five populations where $N > 20$ which were used in the paper by Samiuddin and Asad. First, however, we briefly summarize the methods. All estimators of Y are unbiased.

4.1 THE METHODS

Method 1. The Usual (Textbook) Estimator Under Equal Probability Sampling.

Sampling Scheme. Simple random sampling of size n without replacement from the N units.

Estimator of Y. $\hat{Y} = N\bar{y}$ where $\bar{y}$ is the sample mean.

Variance of Estimator. $Var(\hat{Y}) = N(N-n)\dfrac{S_Y^2}{n}$ where $S_Y^2 = \dfrac{\sum_{i=1}^{N} y_i^2 - \frac{(Y)^2}{N}}{N-1}$.

Method 2. Probability Proportional to Size Sampling With Replacement Using Ranks.

Sampling Scheme. Make n independent selections with replacement. At each selection, the $P(U_i$ is selected$) = p_i = \dfrac{2i}{N(N+1)}$.

Estimator of Y. $\hat{Y}_{HTWR} = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{p_i}$.

Variance of Estimator. $Var(\hat{Y}_{HTWR}) = \frac{1}{n}\sum_{i=1}^{N}\frac{y_i^2}{p_i} - \frac{(Y)^2}{n}$.

Method 3. Rao-Hartley-Cochran (1962) Method Using Ranks.

Sampling Scheme. Divide the population of N units at random into n groups of N/n units each. Independently select one unit from each group with probabilities proportional to the normed measures of size within the group.

Estimator of Y. Let y_{ji} be the sample value if the i^{th} unit of the j^{th} group is selected with normed rank relative to the ranks in that group denoted by p'_{ji}.

$$\hat{Y}_{RHC} = \sum_{j=1}^{n}\frac{y_{ji}}{p'_{ji}} .$$

Variance of Estimator. When each group has N/n units, one can show (See, e.g., Cochran 1977) that

$$Var(\hat{Y}_{RHC}) = \frac{N-n}{N-1}\,Var(\hat{Y}_{HTWR})$$

Method 4. Samiuddin-Asad (1981) Method Using Ranks.

Sampling Scheme. Divide the population of N units into $n+1$ groups. If $P_j = \sum_{\substack{i \text{ in} \\ \text{Group } j}} \pi_i$, the authors recommend forming the groups to satisfy $P_j \leqslant 1 \;\forall\; j$, $P_j + P_{j'} \geqslant 1$ for any two groups j and j', and $P_j = \frac{n}{n+1} \;\forall\; j$, where $j = 1,2,\cdots,n+1$. Actually, the first two inequalities are requirements. Select the j^{th} group with probability $1-P_j$. Omit the selected group. From each of the remaining n groups, independently select one unit with probabilities proportional to the normed measures of size within the group.

Estimator of Y. Let y_j be the value of the unit selected from the j^{th} group and p_j be its measure of size as given in (2.1).

$$\hat{Y}_{HTSA} = \sum_{j=1}^{n}\frac{y_j}{np_j}$$

Variance of Estimator.

$$Var(\hat{Y}_{HTSA}) = \sum_{i=1}^{N}\frac{y_i^2}{\pi_i} - \left[\sum_{j=1}^{n+1}\frac{(1-P_j)}{P_j}Y_j\right]^2 - \sum_{j=1}^{n+1}\left[\frac{2P_j-1}{P_j^2}\right]Y_j^2$$

where $\pi_i = np_i$, $P_j = \sum_{\substack{i \text{ in} \\ \text{Group } j}} \pi_i$, and $Y_j = \sum_{\substack{i \text{ in} \\ \text{Group } j}} y_i$.

Method 5. The method proposed in this paper.

All five estimators are either a special case of the Horvitz-Thompson estimator or a sum of Horvitz-Thompson estimators. For $n > 1$, we note that the sampling variance of Method 3 is always less than the sampling variance of Method 2. Also by Lemma 2, the sampling variance of Method 5 is never greater than the sampling variance of Method 2. Method 3 is not really a case of *πpswor* sampling, though it is an example of unequal probability sampling. Unbiased estimators of the sampling variance are known for Methods 1, 2, and 3.

4.2 THE DATA AND DESCRIPTION OF POPULATIONS

Data from five populations that have appeared in the literature and that will be the basis for our numerical comparisons are given in Table 1. The data in each table are in pairs y/x and are ranked within each population according to the y value from smallest to largest. The y-value is the primary variable of interest and the x-value is an auxiliary variable. Plots of these data are given in Figures 1(a)-5(a). In populations 1, 2, 3, and 4 the data pair with the largest y value was omitted from the original set for all methods to meet the requirement $N = 2kn$. In population 5, the two data pairs with the two largest y values were omitted from the original set to meet this requirement. The corresponding plots of the ranks against the y values are given in Figures 1(b)-5(b).

4.3 VARIANCE AND BIAS COMPARISONS WITH $Var(\hat{Y}_{HTWOR})$

Table 2 gives numerical comparisons of sampling variances of methods 1-4 with the sampling variances of Method 5 for the five populations and various sample sizes. For the populations and sample sizes considered, Method 5 is in every case more efficient than Methods 1 and 2 as expected. For populations 3 and 4, there is very little difference between Methods 2 and 5 because in both cases the points are nearly uniformly scattered on the graphs (see Figures 3(a) and 4(a)), and Method 5 rather successfully creates groups that are quite similar to the overall population. (See next paragraph.) As Table 3 reveals for populations 3 and 4, the $Bias(\hat{Var}(\hat{Y}_{HTWR}))$ relative to $Var(\hat{Y}_{HTWOR})$ is small in every case considered. For populations 3 and 4, Methods 3 and 4 are both more efficient than Method 5. Note that for these populations, the advantage of Method 3 over 5 is not great. For Population 3, the advantage of Method 4 over Method 5 is clear.

Table 1. Data (y/x) for Five Populations

Population 1. Source: Cochran (1977), p. 152. ($N=48$). Population sizes of 48 large United States cities (in 1000's) in 1930 (y_i) and 1920 (x_i).

46/36	53/45	58/50	64/40	79/71	104/93	130/116	260/179
48/23	53/46	60/40	65/46	80/76	105/87	139/136	288/256
50/29	54/36	61/43	67/67	85/94	106/78	142/56	291/243
50/2	57/60	63/37	69/61	86/66	111/30	143/138	317/298
50/43	57/25	63/64	75/48	89/77	113/121	183/172	459/387
52/38	58/44	64/50	77/64	93/74	115/120	232/161	464/381

Population 2. Source: Yates (1949), p. 159. ($N=42$). Population sizes (x_i) of 42 kraals in the Mondora Reserve in Southern Rhodesia and the number of persons absent (y_i) from these kraals.

2/61	7/89	10/82	14/83	18/95	25/124	28/91
3/45	7/47	12/43	15/142	18/125	25/42	21/148
3/31	8/33	12/64	16/65	18/103	26/54	35/85
4/53	9/81	12/75	16/52	19/51	26/132	36/159
5/28	9/57	13/73	16/69	22/86	27/67	41/132
6/30	9/63	14/79	17/43	24/116	27/69	45/96

Population 3. Source: Yates (1949), p. 163. ($N=24$). Measured volumes of timber (y_i) on 24 sample plots and eye estimates (x_i) of corresponding stands (cu.ft. per 1/10 acre).

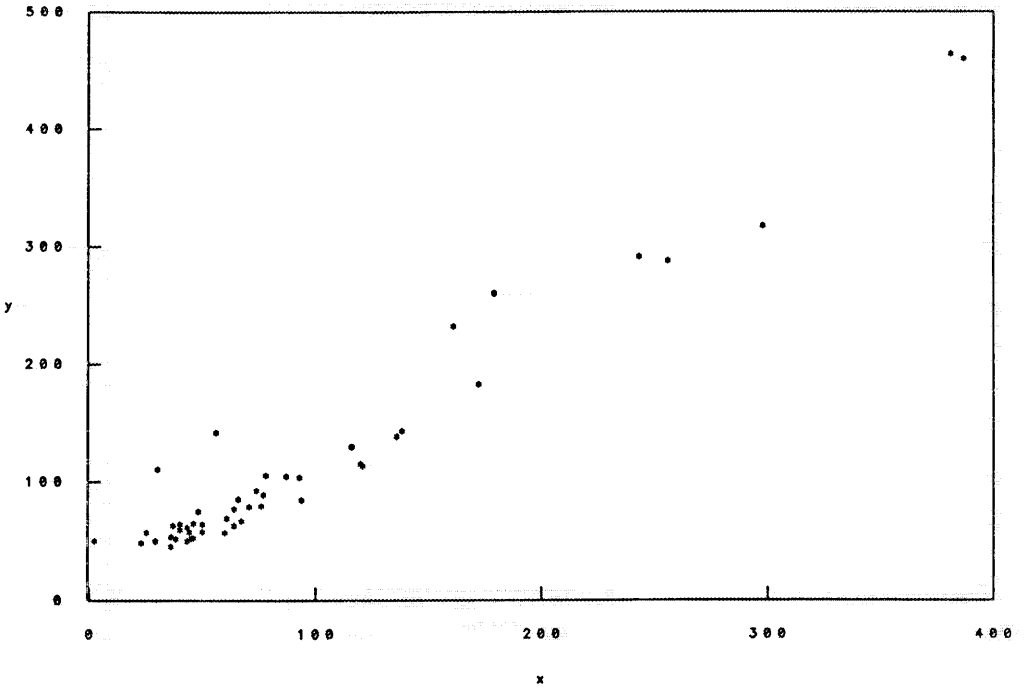
24/207	74/148	100/196	125/65	146/92	154/110	182/152	255/208
47/14	87/110	110/110	126/95	146/110	169/152	195/208	255/167
64/57	91/70	112/128	135/110	153/79	170/102	216/177	261/268

Population 4. Source: Sukhatme and Sukhatme (1970), p. 51. ($N=24$). Values of total cultivated area (x_j) and area under rice (y_i) for 24 villages in Baloda Bazar Tehsil (in acres).

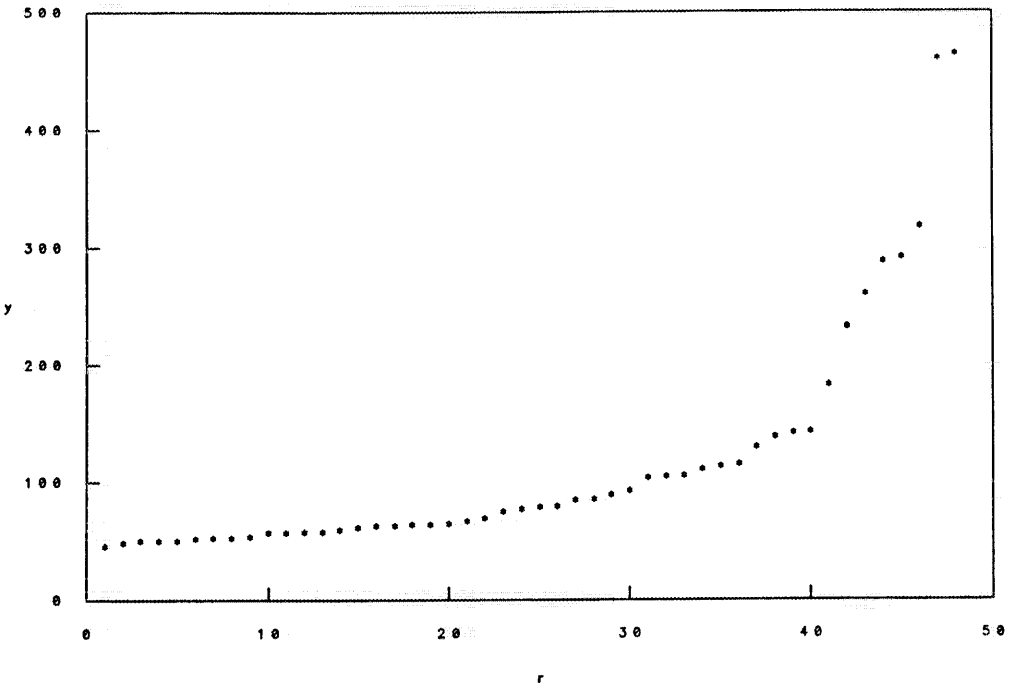
101/120	231/327	392/651	565/671	697/871	785/1042	929/979	1055/1170
107/206	330/515	417/428	568/842	745/1016	810/895	1026/1055	1190/1262
137/162	338/497	516/541	688/1232	768/1346	898/1285	1036/1331	1666/2110

Population 5. Source: Sukhatme and Sukhatme (1970), p. 185 ($N=32$). Values of area under wheat for 1936 (x_i) and 1937 (y_i) for 32 villages in Lucknow Subdivision (India) (in acres).

6/5	62/73	99/109	111/125	141/196	219/236	278/254	355/359
27/45	79/78	100/137	111/125	144/131	221/247	278/254	381/442
52/75	79/62	103/129	112/101	149/163	249/238	289/326	399/427
60/71	85/92	105/78	133/134	179/192	265/255	330/663	498/481

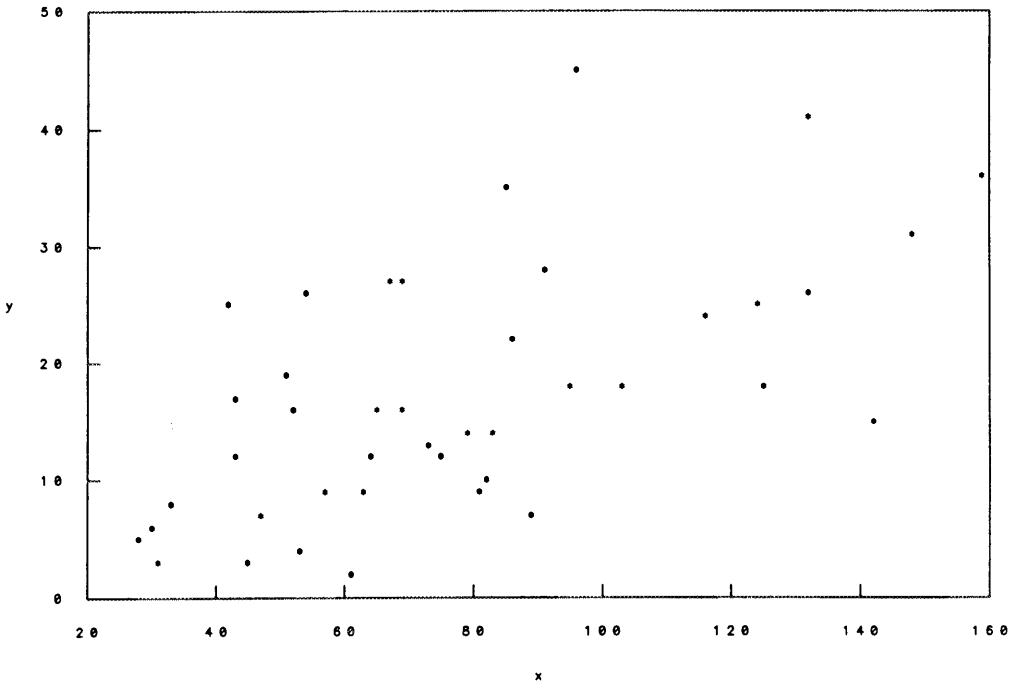


(a) (x, y)

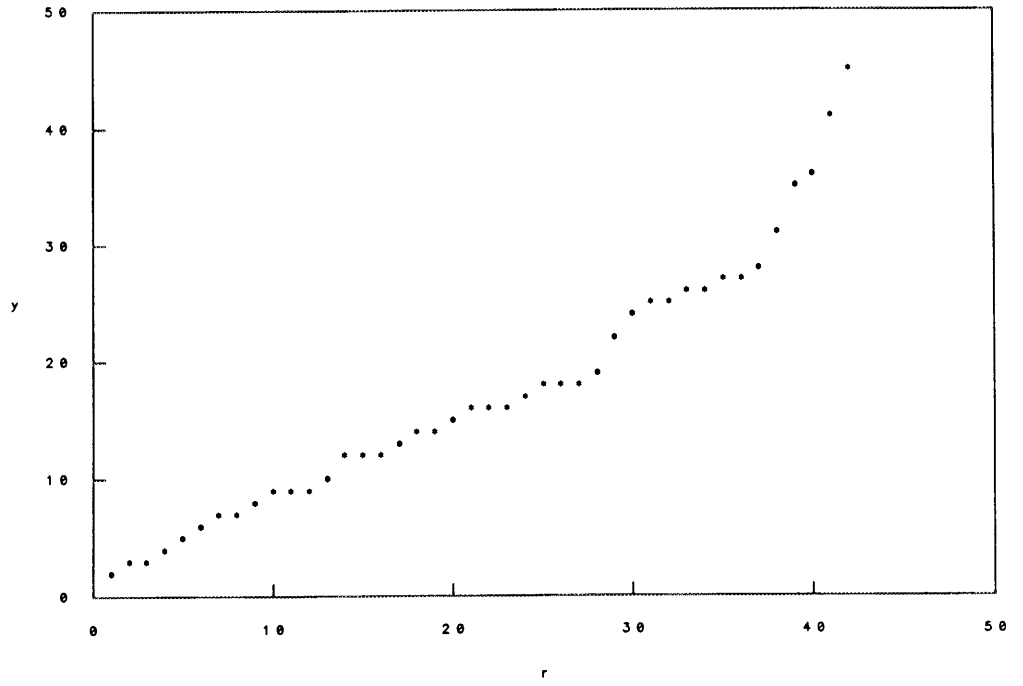


(b) (ranks, y)

Figure 1. Population 1 Plots

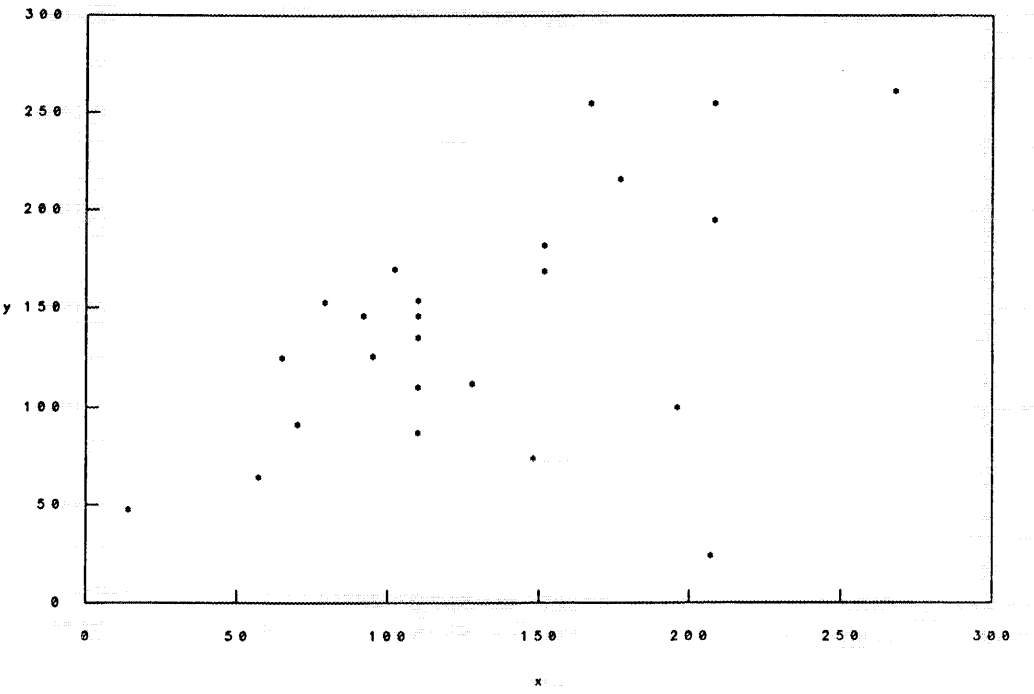


(a) (x, y)

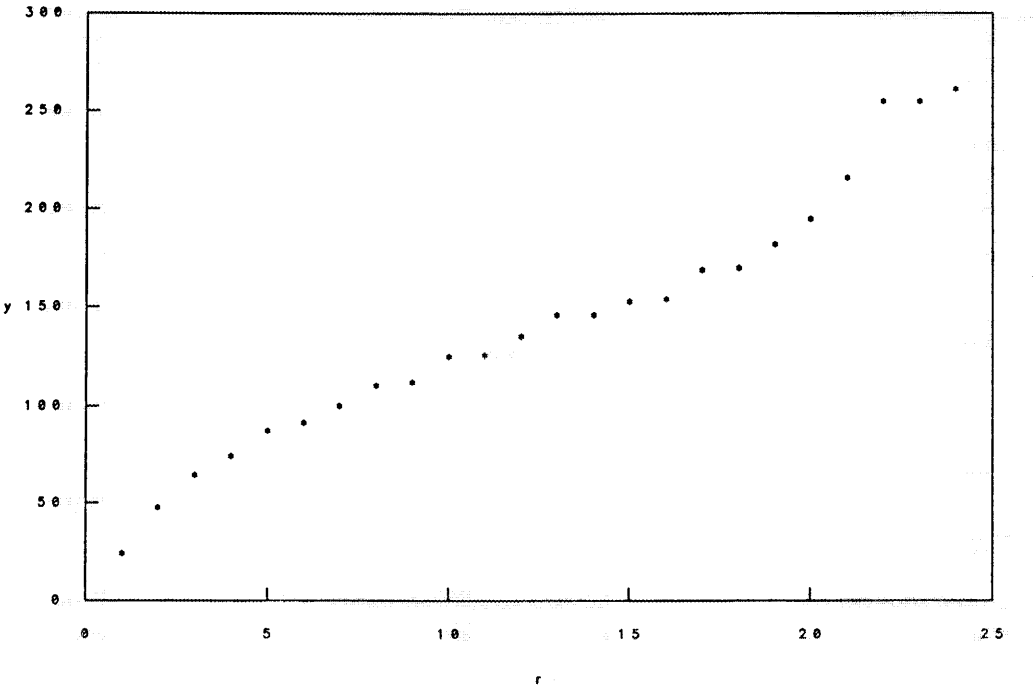


(b) (ranks,y)

Figure 2. Population 2 Plots

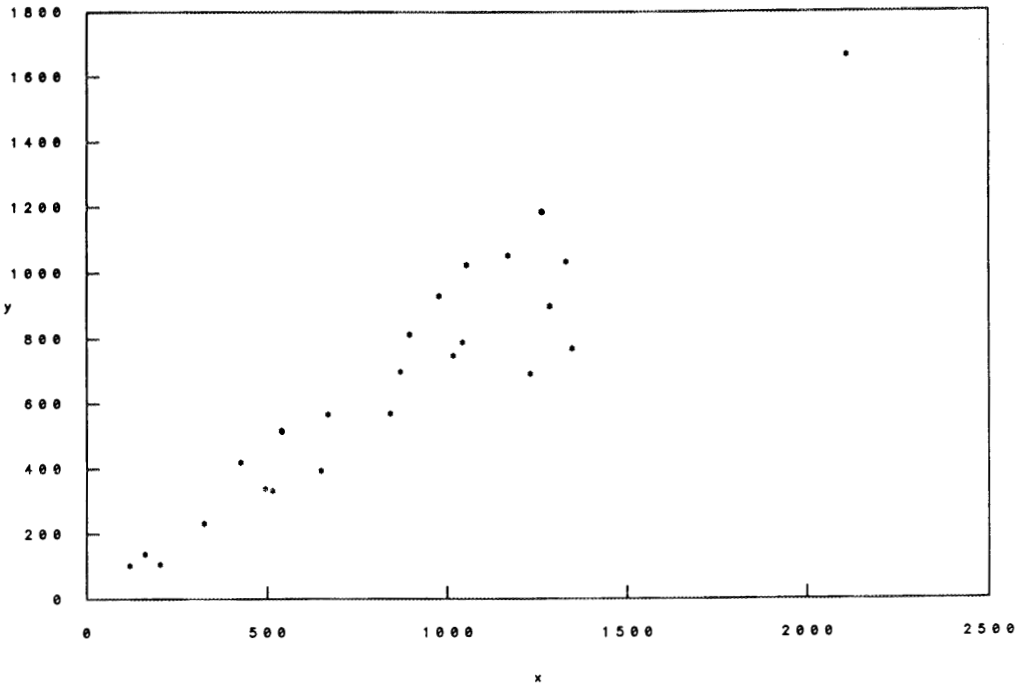


(a) (x, y)

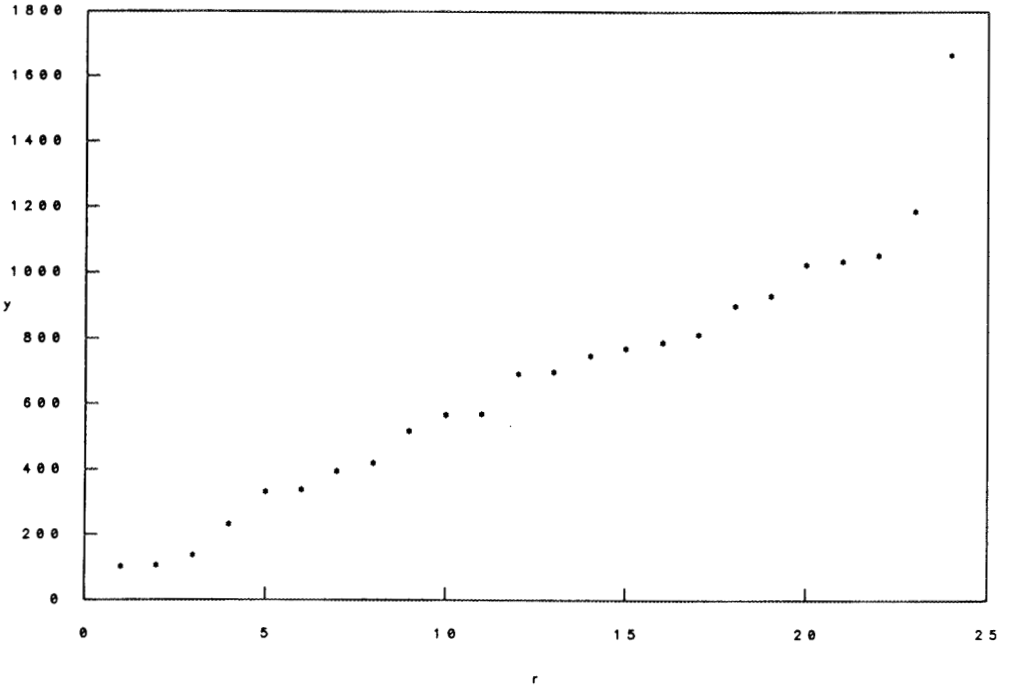


(b) (ranks,y)

Figure 3. Population 3 Plots

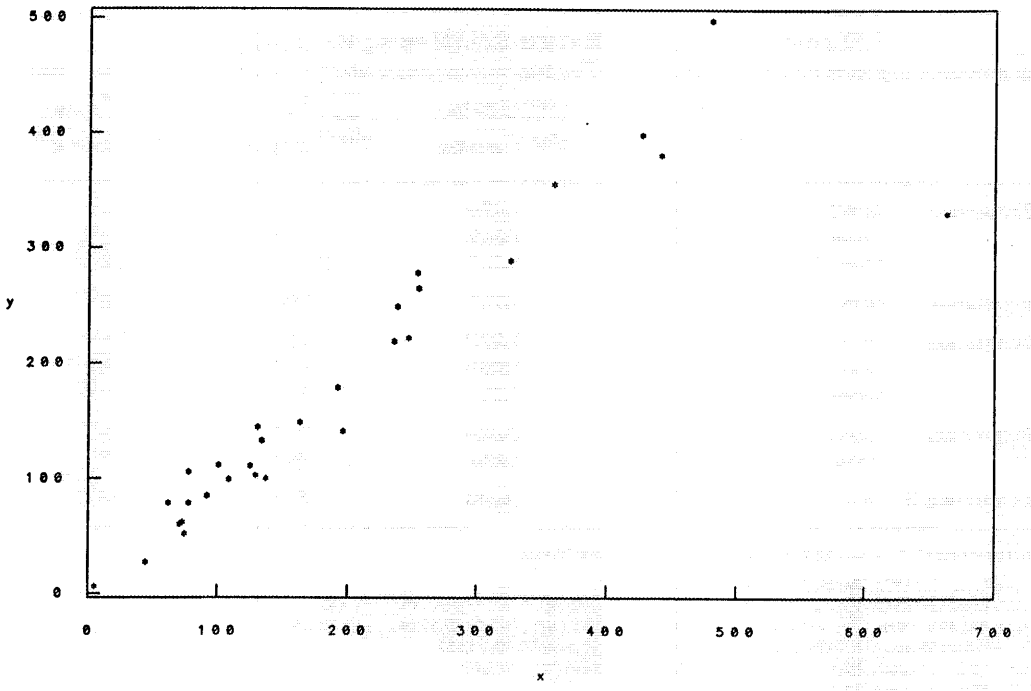


(a) (x ,y)

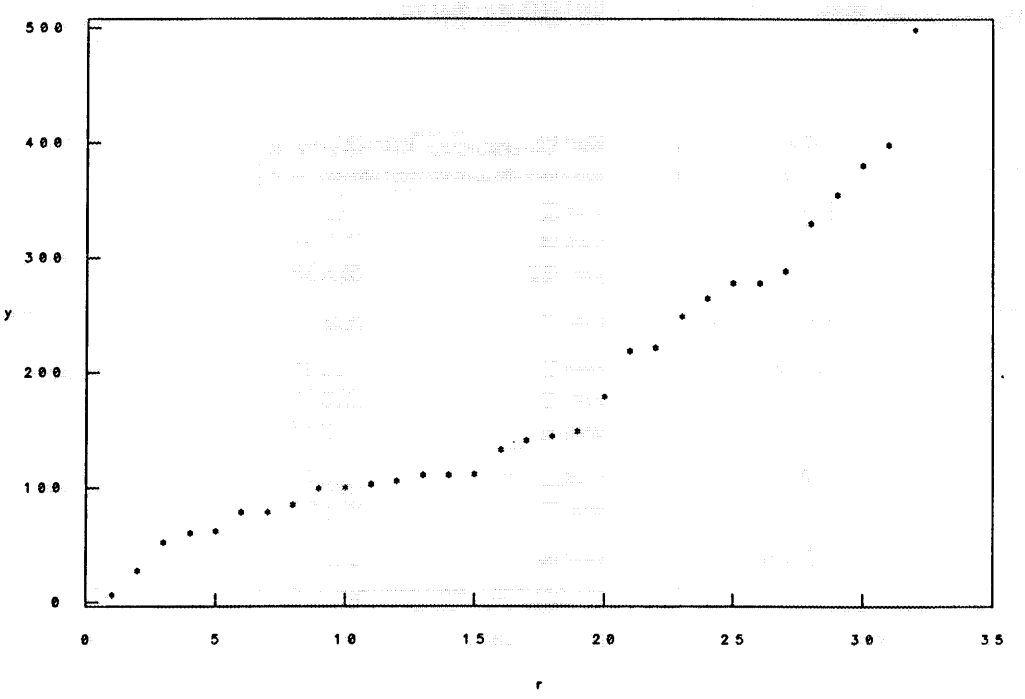


(b) (ranks,y)

Figure 4. Population 4 Plots



(a) (x ,y)



(b) (ranks,y)

Figure 5. Population 5 Plots

Table 2. Variance Ratios Relative to $Var(\hat{Y}_{HTWOR})$

		$\frac{Var(\hat{Y})}{Var(\hat{Y}_{HTWOR})}$	$\frac{Var(\hat{Y}_{HTWR})}{Var(\hat{Y}_{HTWOR})}$	$\frac{Var(\hat{Y}_{RHC})}{Var(\hat{Y}_{HTWOR})}$	$\frac{Var(\hat{Y}_{HTSA})}{Var(\hat{Y}_{HTWOR})}$
Population 1	$n=2$	3.503	1.476	1.444	1.248(1)
	$n=4$	5.335	2.349	2.199	1.589(2)
	$n=12$	6.006	3.233	2.476	1.518(3)
Population 2	$n=3$	31.393	1.787	1.700	1.256(4)
Population 3	$n=2$	4.419	1.022	.978	.799(5)
	$n=3$	4.281	1.038	.947	.664(6)
	$n=6$	3.609	1.021	.799	.408(7)
Population 4	$n=2$	20.726	1.024	.979	.929(8)
	$n=3$	20.508	1.061	.969	.900(9)
Population 5	$n=4$	53.796	5.048	4.559	2.766(10)

Assignments of the Units by Ranks to the $n+1$ groups for Method 4.

(1) (1,2, · · · ,27); (28,29, · · · ,39); (40,41, · · · ,48)
(2) (1,2, · · · ,21); (22,23, · · · ,30); (31,32, · · · ,37); (38,39, · · · ,43); (44,45, · · · ,48)
(3) (47,48); (45,46); (1,2,43,44); (3,4,41,42); (5,6,39,40); (7,8,37,38); (9,10,35,36); (11,12,33,34); (13,14,31,32); (15,16,29,30); (17,18,27,28); (19,24,25,26); (20,21,22,23)
(4) (1,2, · · · ,21); (22,23, · · · ,30); (31,32, · · · ,37); (38,39, · · · ,42)
(5) (1,2, · · · ,14); (15,16, · · · ,20); (21,22,23,24)
(6) (1,2, · · · ,12); (13,14, · · · ,17); (18,19,20,21); (22,23,24)
(7) (1,2, · · · ,9); (10,11,12,13); (14,15,16); (17,18); (19,20); (21,22); (23,24)
(8) (1,2, · · · ,14); (15,16, · · · ,20); (21,22,23,24)
(9) (1,2, · · · ,12); (13,14, · · · ,17); (18,19,20,21); (22,23,24)
(10) (1,2, · · · ,14); (15,16, · · · ,20); (21,22, · · · ,25); (26,27,28,29); (30,31,32)

Table 3. $[Bias(\hat{Var}(\hat{Y}_{HTWR}))]/Var(\hat{Y}_{HTWOR})$

Population 1	$n=2$	.951
	$n=4$	1.799
	$n=12$	2.436
Population 2	$n=3$	1.180
Population 3	$n=2$	.045
	$n=3$	.057
	$n=6$	.025
Population 4	$n=2$	.047
	$n=3$	.092
Population 5	$n=4$	5.397

The numbers in Table 3 confirm that when $Var(\hat{Y}_{HTWOR})$ is much smaller than $Var(\hat{Y}_{HTWR})$, then $\hat{Var}(\hat{Y}_{HTWR})$ is not as good an estimator of $Var(\hat{Y}_{HTWOR})$ in terms of bias as when $Var(\hat{Y}_{HTWOR})$ is just slightly smaller than $Var(\hat{Y}_{HTWR})$. The question arises, when will $\hat{Var}(\hat{Y}_{HTWR})$ be a good estimator of $Var(\hat{Y}_{HTWOR})$ in terms of bias? From the five populations and from the comments following Lemma 3, we see that $\hat{Var}(\hat{Y}_{HTWR})$ will be a good estimator of $Var(\hat{Y}_{HTWOR})$ when the y values are roughly uniformly spaced. For when the y values are roughly uniformly spaced, the means of the n groups under the method of this paper will be essentially the same, as is the case for populations 3 and 4. However, when many of the y values are small and a few are large or when many are large and a few are small, the means of the n groups will differ and the bias will be large. In each of populations 1, 2, and 5, most of the y values are small and relatively few are large. This is easy to see by referring to the (ranks, y) plots for these populations in Figures 1(b)-5(b). In this respect, populations 1, 2, and 5 are typical of many populations in practice. In such case when Method 5 is used, one should consider alternative methods for estimating $Var(\hat{Y}_{HTWOR})$ or alternative methods of sampling such as placing the few large units in a certainty stratum and applying Method 5 to the remaining collection of units.

Is there an advantage in using the ranks as measures of size over other possible auxiliary variables? Table 4 gives comparisons of variances using ranks of the y's ($Var_r(\cdot)$) with variances using the x values ($Var_x(\cdot)$) given in Table 1 for the smallest sample sizes for each population. From Column 3 of Table 4, we see that in 4 out of the 5 comparisons (populations 2, 3, 4, and 5) sampling with replacement using ranks is more efficient than sampling with replacement using the x-variables given in the original sources. From the graphs of population 1 in Figures 1(a) and 1(b), the x's and y's appear to have a stronger linear relation through the origin than do the ranks and y's where the plot seems quadratic and to have a nonzero y-intercept; this may be the reason for the poor showing for the use of the ranks in population 1. However, in every case the sampling without replacement using ranks (Method 5) is more efficient than sampling with replacement using the x variables.

Table 4. Variance Comparisons: Ranks versus Other Measures of Size

Population	n	$\frac{Var_r(\hat{Y}_{HTWR})}{Var_x(\hat{Y}_{HTWR})}$	$\frac{Var_r(\hat{Y}_{HTWOR})}{Var_x(\hat{Y}_{HTWR})}$
1	2	1.281	.868
2	3	.079	.044
3	2	.253	.247
4	2	.516	.504
5	4	.966	.191

5. DISCUSSION

There can be gains in estimation efficiency over *equal probability* and *unequal probability with replacement sampling methods* when one makes use of auxiliary information for *πpswor* sampling methods. However, *πpswor* methods are often complex to execute, particularly when $n > 2$. Also, appropriate auxiliary information may not be available. When the population units can be ranked according to the unknown y's, we demonstrated (Sections 3 and 4) that there can be gains in estimation efficiency with a very simple *πpswor* sampling strategy for *all* sample sizes n where $N = 2kn$.

The use of ranks is common in classical nonparametric statistical methodology. Considering the nonparametric-like nature of sampling theory for a finite population, one wonders why use of the ranks isn't more common in sampling methodology. As we have shown, ranks can be quite useful in *πpswor* sampling, and one may actually have more belief in the accuracy of the ranks as measures of size than the "traditional" x auxiliary variables.

The proposed method requires that $N = 2kn$. If this equality does not hold initially, it can generally be achieved rather easily by taking enough (say c) of the largest units with certainty so that $N-c = 2k(n-c)$. Thus we would select the remaining $n-c$ units for the sample from the remaining $N-c$ units with *πpswor* sampling.

The method of this paper is a clear alternative to the methods of Rao, Hartley, and Cochran (1962) and Samiuddin and Asad (1981). Aside from the results in Section 4, one may compare the methods for forming the groups. Rao, Hartley, and Cochran (1962) form the n groups randomly. As a result, their method is not exact *πpswor* sampling, but it does have an unbiased estimator of variance. Samiuddin and Asad (1981) form $n+1$

groups and eliminate one of them using randomization. The assignment of units to the groups is unclear although Samiuddin and Asad argue for group assignment so that the total π_i's in each group is approximately $n/(n+1)$. In terms of minimizing $Var(\hat{Y}_{HTSA})$, they note (particularly when n is small relative to N) that the groups should be formed so that for the j^{th} group, $y_i/\pi_i = Y_j / \sum_{\substack{i \text{ in} \\ Group\ j}} \pi_i$ for every i in *Group* j. This seems quite time-consuming, and it is not clear how one would accomplish it except by trial-and-error, not to mention that the y_i's and Y_j's would be unknown in practice. Given n, N, and the ranks, the formation of groups for the method (5) of this paper is straightforward and unique.

Let

$$SSB = \frac{N^2}{n^2} \sum_{j=1}^{n} (\mu_{jy} - \mu_y)^2 \ . \tag{5.1}$$

Clearly *SSB* is a measure of variability among the n groups. The quantity *SSB* is of interest because it occurs in two different settings. First, we saw it as the difference between $Var(\hat{Y}_{HTWR})$ and $Var(\hat{Y}_{HTWOR})$; and secondly, we saw it as the (approximate) bias of $\hat{Var}(\hat{Y}_{HTWR})$ as an estimator of $Var(\hat{Y}_{HTWOR})$. In the first case, we want *SSB* to be large if we are arguing that sampling without replacement is more efficient than with replacement; in the latter case, we want *SSB* to be small. The method of stratification used to form the n groups using the ranks suggests that each group tends to have an equal number of y-values on both sides of the overall population mean μ_y, and hence the μ_{jy}'s are expected to be similar in value. Thus, *SSB* should tend to be small rather than large, particularly in populations where the y values are uniformly dispersed as in populations 3 and 4. This is good because to know that we can do no worse using sampling without replacement than sampling with replacement is quite acceptable if we know that we also get the advantage of at most one inclusion in the sample for each population unit, that the *wor* method is simple to execute for all sample sizes, and that a near unbiased variance estimator is possible.

Finally, it should be noted that the method of this paper is one example where we prefer to stratify into heterogeneous groups rather than homogeneous ones.

REFERENCES

[1] Brewer, K. R. W. (1963), "A Model of Systematic Sampling With Unequal Probabilities," *Australian Journal of Statistics*, **5**, pp. 5-13.

[2] Brewer, K. R. W. and Hanif, M. (1983), *Sampling With Unequal Probabilities*, Lecture Notes in Statistics Series, **15**, Springer-Verlag, New York, New York.

[3] Chaudhuri, A. (1975), "A Simple Method of Sampling Without Replacement With Inclusion Probabilities Exactly Proportional to Size," *Metrika*, **22**, pp. 147-152.

[4] Cochran, W. G. (1977), *Sampling Techniques*, 3rd Edition, John Wiley and Sons, New York, New York.

[5] Hansen, M. H. and Hurwitz, W. N. (1943), "On the Theory of Sampling From a Finite Population," *Annals of Mathematical Statistics*, **14**, pp. 333-362.

[6] Horvitz, D. G. and Thompson, D. J. (1952), "A Generalization of Sampling Without Replacement From a Finite Universe," *Journal of the American Statistical Association*, **47**, pp. 663-685.

[7] Madow, W. G. (1949), "On the Theory of Systematic Sampling, II," *Annals of Mathematical Statistics*, **20**, pp. 333-354.

[8] Murthy, M. N. (1957), "Ordered and Unordered Estimators in Sampling Without Replacement," *Sankhya*, **18**, pp. 379-390.

[9] Rao, J. N. K., Hartley, H. O. and Cochran, W. G. (1962), "On a Simple Procedure of Unequal Probability Sampling Without Replacement," *Journal of the Royal Statistical Society, Series B*, **24**, pp. 482-491.

[10] Samiuddin, M. and Asad, H. (1981), "A Simple Procedure of Unequal Probability Sampling," *Biometrika*, **68**, No. 3, pp. 728-731.

[11] Sen, A. R. (1953), "On the Estimate of the Variance in Sampling With Varying Probabilities," *Journal of the Indian Society of Agricultural Statistics*, **5**, pp. 119-127.

[12] Sengupta, S. (1987), "A Comparison Between PPSWR and Chaudhuri's IPPS Procedure," *Metrika* (to appear).

[13] Sukhatme, P. V. and Sukhatme, B. V. (1970), *Sampling Theory of Surveys With Applications*, Iowa State University Press, Ames, Iowa.

[14] Wolter, K. M. (1985), *Introduction to Variance Estimation*, Springer-Verlag, New York, New York.

[15] Yates, F. (1949), *Sampling Methods for Censuses and Surveys*, Hafner Publishing Company, Inc., New York, New York.

[16] Yates, F. and Grundy, P. M. (1953), "Selection Without Replacement From Within Strata With Probability Proportional to Size," *Journal of the Royal Statistical Society, Series B*, **15**, pp. 253-261.

INTERNAL DISTRIBUTION

1. J. J. Beauchamp
2. K. O. Bowman
3. D. J. Downing
4. E. L. Frome
5.-6. R. F.Harbison/ Mathematical Sciences Library
7. T. L. Hebble
8. H. Hwang
9. J. K. Ingersoll
10. E. Leach
11. W. E. Lever
12. F. C. Maienschein
13. S. A. McGuire
14. T. J. Mitchell
15. M. D. Morris
16. G. Ostrouchov
17. R. L. Schmoyer
18. V. R. R. Uppuluri
19.-23. R. C. Ward
24.-28. T. Wright
29. A. Zucker
30. P. W. Dickson (Consultant)
31. G. H. Golub (Consultant)
32. R. M. Haralick (Consultant)
33. D. Steiner (Consultant)
34. Central Research Library
35. K-25 Plant Library
36. ORNL Patent Office
37. Y-12 Technical Library/ Document Reference Section
38. Laboratory Records - RC
39.-40. Laboratory Records Department

EXTERNAL DISTRIBUTION

41. Donald M. Austin, ER-7, Applied Mathematical Sciences, Scientific Computing Staff, Office of Energy Research, Office G-437 Germantown, Washington, DC 20545

42. Barbara A. Bailar, Statistical Research Division, U.S. Bureau of the Census, Washington, DC 20233

43. Leroy Bailey, Statistical Research Division, U.S. Bureau of the Census, 4413 Weldon Drive, Temple Hills, MD 20748

44. Charles B. Bell, Jr., Department of Mathematical Sciences, San Diego State University, College Avenue, San Diego, CA 92182

45. Yvonne M. Bishop, Office of Statistical Standards, Energy Information Administration, Department of Energy, 1000 Independence Avenue, Washington, DC 20585

46. Kenneth R. W. Brewer, Survey Design & Development, Bureau of Agricultural Economics, MacArther House, Lyneham, ACT 2602, AUSTRALIA

47. Michael R. Chernick, Mission Information Systems Division, The Aerospace Corporation, P.O. Box 92957, MS M4-955, Los Angeles, CA 90009

48. Keith R. Eberhardt, Statistical Engineering Division, National Bureau of Standards, Administration Building, Rm. A337, Gaithersburg, MD 20899

49. Nancy Flournoy, Probability and Statistics, National Science Foundation, 1800 G. Street N.W., Washington, DC 20550

50. Wayne A. Fuller, Statistics Department, Iowa State University, 221 Snedecor Hall, Ames, IA 50011

51. V. P. Godambe, Statistics Department, University of Waterloo, University Avenue, Waterloo, Ontario, CANADA N2L 3G1

52. Louis Gordon, Mathematics Department, University of Southern California, DRB 306, Los Angeles, CA 90089-1113

53. Morris H. Hansen, Westat Inc., 1650 Research Blvd., Rockville, MD 20850

54. Graham Kalton, Survey Research Center, University of Michigan, Ann Arbor, MI 48106

55. Nancy Kirkendall, Energy Information Administration, U.S. Department of Energy, 1000 Independence Avenue S.W., Washington, DC 20585

56. Leslie Kish, Institute for Social Research, University of Michigan, Ann Arbor, MI 48106

57. Phillip Kott, Petroleum Marketing Division, Energy Information Administration, E1 43, Rm. 2G043, 1000 Independence Avenue S.W., Washington, DC 20583

58. George Legall, Statistics Department, University of Tennessee, Knoxville, TN 37996

59. Roderick J. Little, Biomathematics Department, University of California at Los Angeles, Los Angeles, CA 90024

60. Robert Latta, Energy Information Administration, U.S. Department of Energy, Mail Stop 1H-053, 1000 Independence Avenue, Washington, DC 20585

61. Ronald Peierls, Applied Mathematics Department, Brookhaven National Laboratory, Upton, New York 11973

62. B. S. Rajput, Mathematics Department, Ayres Hall, University of Tennessee, Knoxville, TN 37996

63. J. N. K Rao, Mathematics and Statistics Department, Carleton University, Ottawa, Ontario, Canada K1S 5B6

64. Richard M. Royall, Biostatistics Department, Johns Hopkins University, 615 N. Wolfe Street, Baltimore, MD 21205

65. Donald B. Rubin, Statistics Department, Harvard University, 1 Oxford Street, Cambridge, MA 02138

66. Jagdish S. Rustagi, Department of Statistics, Ohio State University 1958 Neil Avenue, Columbus, OH 43210

67. A. J. Scott, Department of Mathematics and Statistics, University of Auckland, Auckland, NEW ZEALAND

68. L. R. Shenton, Office of Computing and Information Service, Boyd Graduate Studies Building, University of Georgia, Athens, GA 30602

69. Milton Sobel, Department of Mathematics, University of California at Santa Barbara, Santa Barbara, CA 93017

70. Robert A. Stokes, Energy Systems Department, Pacific Northwest Laboratory, P.O. Box 999, Richland, WA 99352

71. David L. Sylwester, Statistics Department, Stokely Management Center, University of Tennessee, Knoxville, TN 37996

72. How J. Tsao, Eastman Kodak Company, MSD - Kodak Park, Bldg. 56, Rochester, NY 14650

73. Carl Wagner, Mathematics Department, Ayres Hall, University of Tennessee, Knoxville, TN 37996

74. Ray A. Waller, S-1, Statistics, Los Alamos National Laboratory, P.O. Box 1663, Los Alamos, NM 87545

75. William E. Winkler, Statistical Standards, Energy Information Administration, Department of Energy, 1000 Independence Avenue, Washington, DC 22003

76. Kirk M. Wolter, Statistical Research Division, Bureau of the Census, Washington, DC 20233

77. Chien-Fu Jeff Wu, Statistics Department, University of Wisconsin, 1210 W. Dayton Street, Madison, WI 53706

78. Office of Assistant Manager for Energy Research and Development, Department of Energy, Oak Ridge Operations Office, Oak Ridge, Tennessee 37830

79-109. Technical Information Center.