

The Bootstrap and Finite Population Sampling

Synthetic populations are used to study methods for adapting Efron's bootstrap estimation technique to finite population sampling. Of particular interest is the extension of these methods to two-stage cluster sampling. Using simulations based on five artificial populations, two variations of bootstrap estimators and two Taylor series variance estimators for a ratio estimator are compared by mean square errors, stability of the variance estimators, and coverage of the confidence intervals. Generally, there appear to be small differences among the variance estimators, except the bootstrap estimators are somewhat less stable than the Taylor series estimators.

**Data Evaluation and Methods
Research
Series 2, No. 95**

DHHS Publication No. 85-1369

U.S. Department of Health and
Human Services
Public Health Service
National Center for Health Statistics
Hyattsville, Md.
January 1985

Copyright Information

All material appearing in this report is in the public domain and may be reproduced or copied without permission; citation as to source, however, is appreciated.

Suggested Citation

P. J. McCarthy and C. B. Snowden: The bootstrap and finite population sampling. *Vital and Health Statistics. Series 2, No. 95*. DHHS Pub. No. (PHS) 85-1369. Public Health Service. Washington. U.S. Government Printing Office, Jan. 1985.

Library of Congress Cataloging in Publication Data

McCarthy, Philip John

The bootstrap and finite population sampling.

(Data evaluation and methods research. Series 2, no. 95)

Bibliography: p.

Supt. of Docs. no.: HE 20.6209:2/95

1. Health surveys—Statistical methods. 2. Estimation theory. 3. Sampling (Statistics) 4. Nonparametric statistics. I. Snowden, Cecelia B. II. Title.

III. Series: Vital and health statistics. Series 2,

Data evaluation and methods research ; no.

[DNLM: 1. Population Dynamics. 2. Sampling Studies.

3. Probability. W2 A N148vb no. 95]

RA409.U45 362.1'072 84-600282

ISBN 0-8406-0304-5

National Center for Health Statistics

Manning Feinleib, M.D., Dr.P.H., *Director*

Robert A. Israel, *Deputy Director*

Jacob J. Feldman, Ph.D., *Associate Director for Analysis and Epidemiology*

Garrie J. Losee, *Associate Director for Data Processing and Services*

Alvan O. Zarate, Ph.D., *Assistant Director for International Statistics*

E. Earl Bryant, *Associate Director for Interview and Examination Statistics*

Robert L. Quave, *Acting Associate Director for Management*

Gail F. Fisher, Ph.D., *Associate Director for Program Planning, Evaluation, and Coordination*

Monroe G. Sirken, Ph.D., *Associate Director for Research and Methodology*

Peter L. Hurley, *Associate Director for Vital and Health Care Statistics*

Alice Haywood, *Information Officer*

Office of Research and Methodology

Monroe G. Sirken, Ph.D., *Associate Director*

Kenneth W. Harris, *Special Assistant for Program Coordination and Statistical Standards*

Robert J. Casady, Ph.D., *Chief, Statistical Methods Staff*

James T. Massey, Ph.D., *Chief, Survey Design Staff*

Acknowledgments

The senior author gratefully acknowledges the financial and computing support of the National Center for Health Statistics, where he spent a sabbatical leave during the spring semester of 1983 when much of the preparation for this report was performed. Special thanks are owed to Drs. Monroe Sirken and Robert Casady who arranged this visit, to Dr. Robert Casady who acted as a sounding board for the ideas and for the plans for simulations, to Dr. Jai Choi who was most helpful in all respects, and to all individuals in the Office of Research and Methodology who made my stay so pleasant.

Contents

Introduction	1
Description of methods	2
Populations used in the investigation	5
Variance estimation	7
The estimation of bias and an adjusted ratio estimator	10
Coverage of confidence intervals	12
Nonparametric confidence intervals for R	15
Two-stage sampling	17
Summary	22
References	23

List of text figures

1. Diagram of the basic population	5
2. Diagram of the $y^{1.5}$ population	6
3. Diagram of the $y + 1.5x - 0.1x^2$ population	6
4. Scatter diagram of 15 clusters and 12 elements for population A	19
5. Scatter diagram of 10 clusters and 18 elements for population B	19

List of text tables

A. Characteristics of populations used in the investigation	6
B. Underpercent and overpercent estimates of $MSE(\hat{R})$ by population and estimated parameters	8
C. Mean square errors of mean square error estimates by population and estimated parameters	9
D. Estimation of bias of \hat{R} using the bootstrap	10
E. Bias of the adjusted ratio estimator	11
F. Percent of 500 samples for which 2-tailed t -statistic equals or exceeds the tabular value	13
G. Comparison of effects of variance estimators from table F	14
H. Comparison of effects of different populations from table F	14
J. Comparison of effects of different sample sizes from table F	14
K. Coverage and average width of nonparametric 95 percent confidence intervals	16
L. Location of R with respect to specified percentile intervals of \hat{R}^* with BWR	16
M. Population used in the investigation of clustered populations and samples	20
N. Underpercent and overpercent estimates of $MSE(\hat{R})$ for clustered populations and samples	20
O. Mean square errors of mean square error estimates for clustered populations and samples	20
P. Comparison of actual and estimated bias of \hat{R} by each population using the bootstrap for cluster sampling	20
Q. Percent of 500 samples for which 2-tailed t -statistic equals or exceeds the tabular value	21

The Bootstrap and Finite Population Sampling

by Philip J. McCarthy, Ph.D., Cornell University, and
Cecelia B. Snowden, M.A., Office of Research and
Methodology, National Center for Health Statistics

Introduction

Efron¹ introduced the bootstrap as a nonparametric method for providing answers to various statistical problems by "the substitution of raw computing power for theoretical analysis."² The following brief description of the bootstrap is freely adapted from references 1-3.

Suppose the data points (univariate or multivariate) y_1, y_2, \dots, y_n are independent observations from a distribution function F , and that one wishes to estimate and study the properties of some parameter of F , say $\theta(F)$. The distribution function F can be estimated by the empirical probability function

$$\hat{F}: \text{mass } \frac{1}{n} \text{ on each observed data point } y_i, i = 1, 2, \dots, n$$

A bootstrap sample is obtained by making n random draws

with replacement from $\{y_1, y_2, \dots, y_n\}$, namely $\{Y_1^*, Y_2^*, \dots, Y_n^*\}$. An estimate of the parameter $\theta^*(Y_1^*, Y_2^*, \dots, Y_n^*)$ is obtained from this bootstrap sample. This obviously can be repeated any desired number of times, say B , leading to B independent estimates of the parameter

$$\theta_1^*(\hat{F}), \theta_2^*(\hat{F}), \dots, \theta_B^*(\hat{F})$$

and these values can be used in a variety of ways to study the distribution of $\hat{\theta}^*(\hat{F})$; for example, to estimate its variance, standard deviation, bias, and percentiles. References to a variety of applications are given in Efron and Gong.² Asymptotic results for the bootstrap have been published by Bickel and Freedman.⁴ Further applications are described by Diaconis and Efron.⁵

Description of methods

Design-based, finite population sampling theory, for simple random sampling without replacement, may be defined as follows. Given a finite population of N elements, a sample of n is selected one at a time by choosing elements in such a fashion that, at each stage of the selection, each of the remaining (undrawn) elements has an equal chance of being selected. This means that selections are not independent of one another. As a matter of fact, the correlation between the variate value of the element selected at the i th draw and that selected at the j th draw is $-1/(N-1)$. Such sampling designs frequently are characterized by the sampling fraction, $f = n/N$.

When one attempts to apply Efron's bootstrap to finite population samples, difficulties are encountered. Perhaps the easiest way of demonstrating this is by the following observations:

1. If the sample data points y_1, y_2, \dots, y_n are n independent observations from a distribution F , then any bootstrap sample is a set of observations that might have been obtained as a sample of n independent observations from F .
2. If the sample data points y_1, y_2, \dots, y_n refer to n elements drawn from a finite population of N elements, there is no way of selecting n elements from the sample that will produce a sample that might have been drawn from the original population; for example,
 - a. If the n elements are selected with replacement from the sample, repetitions of elements can occur in the bootstrap sample. This, by definition, would not be possible in an original sample from the finite population.
 - b. If the n elements are selected without replacement, then one simply obtains the original sample back again.

In searching for ways to adapt the bootstrap idea to finite population sampling, several possibilities were considered. One method is to use the sample to create an artificial population from which repeated simple random samples could be drawn without replacement. For example, if $N = kn$, that is, the sampling fraction $f = n/N = 1/k$, then each of the sample elements can be replicated k times to create a finite population of $kn = N$ elements. If samples of n elements are drawn without replacement from this artificial population, then the sampling fraction $n/N = n/kn$ is still $1/k$. In the case of a linear estimator, such as $\bar{y} = \bar{Y}$, we obtain the following results for this procedure, where \bar{y}^* is the mean of a sample of n drawn from the artificial

population:

$$\bar{y}^* = \frac{\sum_{i=1}^n y_i^*}{n} \quad (1)$$

$$E(\bar{y}^* | \text{original sample of } n) = \bar{y}$$

$$\begin{aligned} V^*(\bar{y}^* | \text{original sample of } n) &= \frac{kn - n}{kn} \frac{1}{n} \frac{k \sum_{i=1}^n (y_i - \bar{y})^2}{kn - 1} \\ &= (1 - f) \frac{1}{n} \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n - 1/k} \end{aligned} \quad (2)$$

where the estimated variance of the original sample mean \bar{y} is given by

$$\begin{aligned} \hat{V}(\bar{y}) &= \frac{N - n}{N} \frac{1}{n} \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n - 1} \\ &= (1 - f) \frac{1}{n} \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n - 1} \end{aligned} \quad (3)$$

Thus the variance of the sample means \bar{y}^* of repeated samples drawn from the artificial population must be multiplied by the factor $(n - 1/k)/(n - 1)$ to obtain the variance one would hope to get, $\hat{V}(\bar{y})$. For example, if $n = 5$ and $k = 2$, the factor is 1.125, while if $n = 5$ and $k = 10$, it is 1.225. Obviously if n is large, its value cannot be far from 1. This is the approach used by Bickel and Freedman⁶ to obtain asymptotic results for stratified simple random sampling without replacement, a method suggested earlier by Gross.⁷ This method will be labeled the without-replacement bootstrap (BWO).

This analysis also can be made if f is not 1 divided by an integer. Suppose $f = k_1/k_2$, where k_1 and k_2 are integers, $k_1 < k_2$. The procedure now is to replicate each sample element k_2 times and select, without replacement, a sample of $k_1 n$ ele-

ments. As before,

$$\bar{y}^* = \frac{\sum_{i=1}^{k_1 n} y_i^*}{k_1 n} \quad (4)$$

$$E(\bar{y}^* | \text{original sample of } n) = \bar{y}$$

$$V^*(\bar{y}^* | \text{original sample of } n) = \frac{nk_2 - nk_1}{nk_2} \frac{1}{k_1 n} \frac{k_2 \sum_{i=1}^n (y_i - \bar{y})^2}{nk_2 - 1}$$

$$= (1 - f) \frac{1}{n} \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{k_1(n - 1/k_2)} \quad (5)$$

Now,

$$\hat{V}(\bar{y}) = \frac{k_1(k_2 n - 1)}{k_2(n - 1)} V^*(\bar{y}^* | \text{original sample of } n)$$

This factor might well be of a larger magnitude than one would like to introduce into variance computations.

An alternative method that may work in some circumstances, and that does not require any correction factor, was suggested by Bickel and Freedman.⁶ Suppose $N = kn + r$, $1 \leq r \leq n - 1$. Construct two artificial populations in the following manner:

<i>Population</i>	<i>Composition of population</i>
1	Replicate each sample element k times, and select samples of n without replacement.
2	Replicate each sample element $k + 1$ times, and select samples of n without replacement.

The variances of the sample means for these two populations are as follows:

<i>Population</i>	<i>Variance</i>
1	$V_1^*(\bar{y}^* \text{original sample of } n)$ $= \frac{kn - n}{kn} \frac{1}{n} \frac{k \sum_{i=1}^n (y_i - \bar{y})^2}{kn - 1}$
2	$V_2^*(\bar{y}^* \text{original sample of } n)$ $= \frac{(k + 1)n - n}{(k + 1)n} \frac{1}{n} \frac{(k + 1) \sum_{i=1}^n (y_i - \bar{y})^2}{(k + 1)n - 1}$

Select in the bootstrap samples population 1 with probability α and population 2 with probability $1 - \alpha$, where α is chosen so

that

$$0 < \alpha < 1$$

and

$$\alpha V_1^*(\bar{y}^* | \text{original sample of } n) + (1 - \alpha) V_2^*(\bar{y}^* | \text{original sample of } n)$$

$$= \frac{N - n}{N} \frac{1}{n} \frac{\sum (y_i - \bar{y})^2}{n - 1}$$

That is, so that

$$\alpha \frac{k - 1}{kn - 1} + (1 - \alpha) \frac{k}{(k + 1)n - 1} = \left(1 - \frac{n}{N}\right) \frac{1}{n - 1} \quad (6)$$

This procedure is not feasible if both $(k - 1)/(kn - 1)$ and $k/[(k + 1)n - 1]$ are less than $(1 - n/N)/(n - 1)$. For example, if $k = 2$, $n = 5$, and $N = 12$,

$$\frac{k - 1}{kn - 1} = 0.1111$$

$$\frac{k}{(k + 1)n - 1} = 0.1429$$

and

$$\frac{1 - n/N}{n - 1} = 0.1458$$

It also is not feasible if $k = 1$. On the other hand, if $k = 2$, $n = 8$, and $N = 20$,

$$0.0667\alpha + (1 - \alpha)0.0870 = 0.0857$$

and $\alpha = 0.064$.

The second procedure considered for bootstrapping finite population samples was obtained from variance considerations. If we have a sample of n drawn without replacement from a finite population of N elements, then $\hat{V}(\bar{y})$ is given by equation (3). Suppose now that a sample of n^* elements is drawn with replacement from the n sample elements and we define

$$\bar{y}^* = \frac{\sum_{i=1}^{n^*} y_i^*}{n^*}$$

Then

$$E(\bar{y}^* | \text{original sample}) = \bar{y}$$

and

$$V^*(\bar{y}^* | \text{original sample}) = \frac{1}{n^*} \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}$$

If we now choose

$$n^* = \frac{n-1}{1-f} = \frac{k}{k-1}(n-1) \quad (7)$$

$V(\bar{y}^* | \text{original sample})$ will be equal to $\hat{V}(\bar{y})$. This method will be labeled the with-replacement bootstrap (BWR).

Rao and Wu⁸ have suggested a different procedure for achieving the same goal as BWR. They would transform y_i^* to

$$\tilde{y}_i = \bar{y} + m^{1/2}(n-1)^{-1/2}(1-f)^{1/2}(y_i^* - \bar{y})$$

where m is the size of the bootstrap sample drawn by BWR from the original sample of n . It is easy to show that $E(\tilde{y} | \text{original sample}) = \bar{y}$ and that $V(\tilde{y} | \text{original sample})$ is equal to $\hat{V}(\bar{y})$ as given by equation (3). This has the advantage of avoiding the rounding problem that occurs when n^* is not an integer. This approach did not come to our attention until after all work on this report had been completed. Otherwise it would have been included among the simulations described.

It is not immediately apparent which of the two procedures, BWO or BWR, should be preferred. The first seems intuitively appealing to some, and the second appeals to others. The second

seems more closely related to Efron's bootstrap, and the first seems more related to traditional finite population sampling theory. In a certain limiting sense they are almost identical. Thus, if k becomes very large, that is, if f becomes small, the first procedure becomes essentially BWR from the sample because each element is replicated a large number of times. We will later consider two-stage sampling; there the first procedure becomes very clumsy. We will attempt to provide comparisons of the two procedures, except for two-stage sampling.

This investigation will focus on the behavior of the ratio estimator

$$\hat{R} = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i} \quad (8)$$

where

$$R = \frac{\sum_{i=1}^N y_i}{\sum_{i=1}^N x_i}$$

A variety of artificial populations and sample sizes will be used in our study.

Populations used in the investigation

Because the ratio estimator has optimum properties when the relationship between y and x is linear through the origin with variance of y for fixed x proportional to x (Cochran,⁹ p. 158), we first constructed a finite population having approximately these properties. The variable x was assumed to have a χ^2 distribution with nine degrees of freedom. Then, for a fixed value of x , y was defined to be

$$y = x + N(0, x)$$

In the infinite superpopulation defined by these relationships we have

$$\left. \begin{aligned} E(x) &= 9 \\ V(x) &= 18 \\ E(y|x) &= x \\ E(y) &= 9 \\ V(y|x) &= x \\ V(y) &= 27 \\ \text{Cov}(x, y) &= 18 \\ \rho_{xy} &= 0.8165 \end{aligned} \right\} \quad (9)$$

To obtain a finite population of $N = 100$ elements, we selected 100 independent x values. For each of these an observation was selected from $N(0, x)$, which, when added to the observed x value, produced an associated y value. The population actually obtained is shown in figure 1, and some of its characteristics are given in table A. Because our other populations were obtained by performing various transformations on this first population, we have chosen to refer to it as the basic population.

The second population was obtained from the basic population simply by increasing each y value by 20, leaving the x values unchanged. The line of the relationship between y and x no longer is close to the origin. Some of its characteristics are given in table A. It will be referred to as $y + 20$.

The third population was obtained from basic by replacing

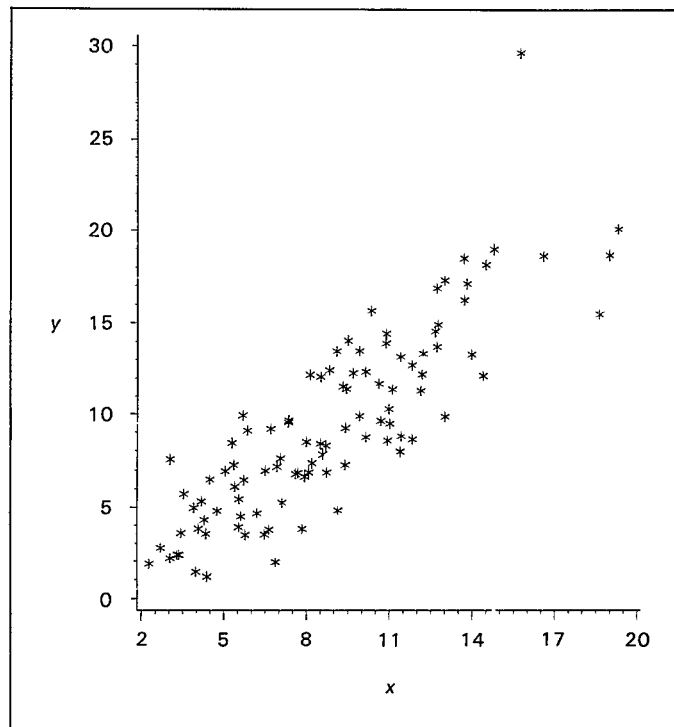


Figure 1. Diagram of the basic population

each y value by $y^{1.5}$, the x values remaining unchanged. The intent was to introduce some curvature into the relationship between y and x . This was not completely successful, as is evident from figure 2 and the correlation shown in table A. It will be designated as $y^{1.5}$.

The fourth population was obtained from the third by increasing each y value by 20. Again, its characteristics are shown in table A.

The fifth population was obtained in the following manner. For each point in the basic population, the deviation of y from x was obtained. This deviation was then added to the quadratic term $2.5x - 0.1x^2$ to obtain a new y value. That is, the new y values are given by $y - x + 2.5x - 0.1x^2 = y + 1.5x - 0.1x^2$. The x values again were left unchanged. This population is shown in figure 3 and some of its characteristics are given in table A. Note that this population exhibits distinct curvature and that the correlation between x and y has been reduced to 0.62. This population will be denoted as $y + 1.5x - 0.1x^2$.

Table A. Characteristics of populations used in the investigation

[See text for explanation of symbols]

<i>Population</i>	<i>n</i>	ρ_{xy}	<i>R</i>	\bar{R}	$\hat{V}(\bar{R})$	$\frac{\text{Bias}}{\sqrt{\hat{V}(\bar{R})}}$
						<i>Percent</i>
Basic	10	0.8488	1.0609	1.0607	0.0084	-0.2
	20			1.0602	0.0038	-1.1
	50			1.0605	0.0009	-1.3
$y + 20$	10	0.8488	3.3168	3.3557	0.0962	12.6
	20			3.3373	0.0418	10.0
	50			3.3193	0.0097	2.6
$y^{1.5}$	10	0.8185	3.6006	3.5803	0.3071	-3.7
	20			3.5865	0.1349	-3.8
	50			3.5957	0.0337	-2.7
$y^{1.5} + 20$	10	0.8185	5.8565	5.8668	0.2461	2.1
	20			5.8657	0.1135	2.7
	50			5.8582	0.0278	1.0
$y + 1.5x - 0.1x^2$	10	0.6198	1.5109	1.5242	0.0229	8.8
	20			1.5174	0.0104	6.4
	50			1.5122	0.0028	2.5

NOTE: Estimates (last 3 columns) are based on 5,000 samples.

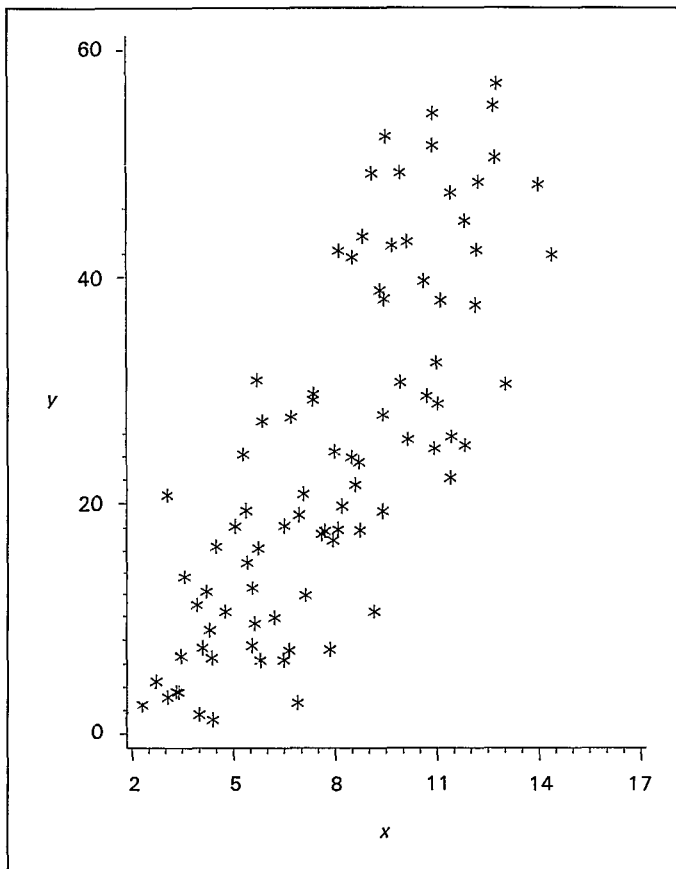


Figure 2. Diagram of the $y^{1.5}$ population

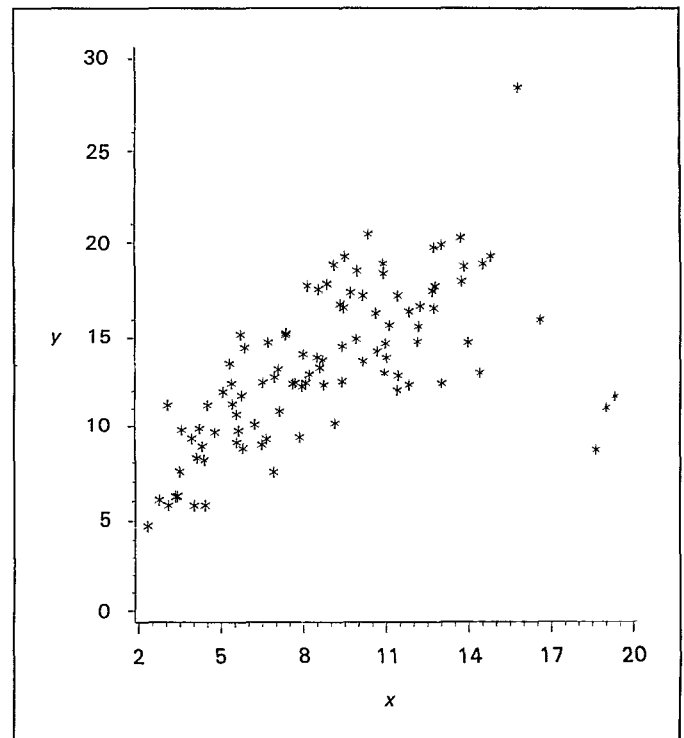


Figure 3. Diagram of the $y + 1.5x - 0.1x^2$ population

Variance estimation

After a sample has been selected from a finite population and an estimate of a population parameter has been computed, one ordinarily estimates the sampling variability of the estimate. Because our concern in this investigation is with the ratio estimator, we have used as our basis of comparison the two standard Taylor series estimators of variance:

$$\hat{V}_{T\bar{x}}(\hat{R}) = \frac{N-n}{N} \frac{1}{n} \frac{1}{\bar{x}^2} \frac{\sum_{i=1}^n (y_i - \hat{R}x_i)^2}{n-1} \quad (10)$$

and

$$\hat{V}_{T\bar{X}}(\hat{R}) = \frac{N-n}{N} \frac{1}{n} \frac{1}{\bar{X}^2} \frac{\sum_{i=1}^n (y_i - \hat{R}x_i)^2}{n-1} \quad (11)$$

where \bar{x} is the sample mean of the x values and \bar{X} is the population mean of these values. The properties of these two estimators of variance are discussed in Cochran⁹ (section 6.4).

For BWO, for each of 500 samples we drew 100 bootstrap samples, thus giving $\hat{R}_1^*, \hat{R}_2^*, \dots, \hat{R}_{100}^*$. From these we computed

$$\bar{\hat{R}}_{BWO}^* = \frac{\sum_{i=1}^{100} \hat{R}_i^*}{100} \quad (12)$$

and

$$\hat{V}_{BWO}(\hat{R}) = \frac{\sum_{i=1}^{100} (\hat{R}_i^* - \bar{\hat{R}}_{BWO}^*)^2}{99} \frac{n-1/k}{n-1} \quad (13)$$

Thus $\hat{V}_{BWO}(\hat{R})$ is the BWO estimate of the variance of \hat{R} . We chose 100 bootstrap samples on the basis of Efron's recommendation. Efron¹⁰ (p. 317) states, "For only estimating $\hat{\sigma}_{boot}$, $N = 100$ performs quite well in most examples." We also note that Efron³ suggests that $(\bar{\hat{R}}_{BWO}^* - \hat{R})$ provides an estimate of the ratio bias of \hat{R} , that is, of $\hat{R} - R$. Thus we can compute directly an estimate of the mean square error (MSE) of \hat{R} as

$$M\hat{S}E_{BWO}(\hat{R}) = \hat{V}_{BWO}(\hat{R}) + (\bar{\hat{R}}_{BWO}^* - \hat{R})^2$$

The ordinary single sample approach, in conjunction with Taylor series estimators of variance, provides no estimate of bias. This estimate of bias will be discussed in greater detail later in this report.

Finally, these various quantities were summarized over 500 samples drawn from the original population, giving rise to

$$\begin{aligned} & \bar{\hat{R}} \\ & \hat{V}(\hat{R}) \\ & \bar{\hat{V}}_{T\bar{x}}(\hat{R}) \\ & \hat{V}(\hat{V}_{T\bar{x}}(\hat{R})) \\ & \bar{\hat{V}}_{T\bar{X}}(\hat{R}) \\ & \hat{V}(\hat{V}_{T\bar{X}}(\hat{R})) \\ & \bar{\hat{R}}_{BWO}^* \\ & (\bar{\hat{R}}_{BWO}^* - \bar{\hat{R}}) \\ & \hat{V}(\bar{\hat{R}}_{BWO}^*) \\ & \bar{\hat{V}}_{BWO}(\hat{R}) \\ & \hat{V}(\hat{V}_{BWO}(\hat{R})) \\ & \overline{M\hat{S}E}_{BWO}(\hat{R}) \\ & \hat{V}(M\hat{S}E_{BWO}(\hat{R})) \end{aligned}$$

Our comparisons of the various procedures will be made on the basis of this average behavior.

Through an oversight in preparing the program, the factor $(n-1/k)/(n-1)$ was not used on every individual $\hat{V}_{BWO}(\hat{R})$. Instead, it was introduced after $\bar{\hat{V}}_{BWO}(\hat{R})$ had been computed. This causes no difficulty; however, it does create a problem with respect to $\overline{M\hat{S}E}_{BWO}(\hat{R})$. Inserting the factor $(n-1/k)/(n-1)$ after the average mean square error has been computed results in this factor multiplying the average squared bias as well as the average variance. Because the squared bias is usually of little consequence in computing MSE's, we do not feel that this error will seriously affect our conclusions.

Using a different set of 500 samples drawn from the original population, this entire process was repeated for the BWR procedure. We thus obtained exactly the same quantities just defined for BWO: \hat{R}_{BWR}^* , $\hat{V}_{BWR}^*(\hat{R})$, and so forth. Note, however, that no correction is required in computing $\hat{V}_{BWR}^*(\hat{R})$.

These computations were performed for each of the five populations described in table A and for each of the sample sizes $n = 10, 20,$ and 50 . For these sample sizes we have

n	k	n_1^* for BWR
10	10	10
20	5	23.75 = 24
50	2	98

Finally, because theory does not generally provide the exact values of $E(\hat{R})$ and $MSE(\hat{R})$, we approximated these quantities by drawing 5,000 samples from each population for each sample size. These values are given in table A.

The results of these simulations, as far as variance estimation is concerned, are presented in table B. We give the percent errors in estimating $MSE(\hat{R})$ for the various combinations of sample size, population, and variance estimator. For example, for the basic population and samples of size 10

$$\frac{\overline{\hat{V}_{Tx}(\hat{R})} - MSE(\hat{R})}{MSE(\hat{R})} \times 100 = -2.6 \text{ percent}$$

Table B. Underpercent and overpercent estimates of $MSE(\hat{R})$ by population and estimated parameters

[See text for explanation of symbols]

n and population	$\overline{\hat{V}_{Tx}(\hat{R})}$		$\overline{\hat{V}_{Tx}(\hat{R})}$		$\overline{\hat{V}_{BWR}(\hat{R})}$		$\overline{\hat{V}_{BWO}(\hat{R})}$		$\overline{M\hat{S}E_{BWR}(\hat{R})}$		$\overline{M\hat{S}E_{BWO}(\hat{R})}$	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
n = 10												
Basic	-2.6	-6.6	-3.0	-5.5	0.2	-4.2	1.3	-3.1				
y + 20	-2.1	-3.1	-10.2	-10.3	8.8	8.0	11.9	10.8				
y ^{1.5}	-4.5	-19.2	0.7	-17.7	-3.9	-18.5	2.3	-17.3				
y ^{1.5} + 20	-18.2	-9.5	-19.1	-8.4	-15.9	-8.3	-14.7	-6.7				
y + 1.5x - 0.1x ²	-8.3	-7.3	-7.2	-4.2	-6.2	-3.2	-4.0	-1.2				
Mean	-7.1	-9.1	-7.8	-9.2	-3.4	-5.2	-1.6	-3.5				
n = 20												
Basic	-3.5	1.6	-2.9	+2.3	-3.6	3.5	-2.4	4.5				
y + 20	-5.4	-1.9	-7.8	-6.5	-2.5	3.2	-0.9	5.0				
y ^{1.5}	-8.0	-2.6	-6.7	-1.0	-8.4	-1.8	-7.3	-0.8				
y ^{1.5} + 20	-6.3	-4.1	-3.9	-3.1	-7.3	-3.0	-6.4	-2.0				
y + 1.5x - 0.1x ²	-6.2	-2.4	-5.4	-0.1	-6.7	-1.0	-5.5	0.4				
Mean	-5.9	-1.9	-5.3	-1.7	-5.7	0.2	-4.5	1.4				
n = 50												
Basic	-2.1	-0.6	-2.5	-1.1	-2.0	-0.0	-1.0	-0.1				
y + 20	1.6	1.8	0.8	1.0	2.1	3.2	3.3	4.4				
y ^{1.5}	-3.5	-3.1	-3.0	-2.4	-4.3	-2.3	-3.2	-1.3				
y ^{1.5} + 20	-3.2	1.9	-2.5	-1.7	-4.7	-1.5	-3.6	-0.6				
y + 1.5x - 0.1x ²	-5.0	-3.4	-5.0	-2.9	-4.8	-3.4	-3.8	-2.3				
Mean	-2.4	-0.7	-2.4	-1.4	-2.7	-0.8	-1.7	0.0				

NOTES: Values of $MSE(\hat{R})$ are based on 5,000 samples; $\overline{\hat{V}}$ values are based on 500 samples; BWR and BWO values are based on 100 bootstrap samples.

On any line in this table, the entries in column (1) are based on 1 set of 500 samples; the entries in column (2) are based on an independent set of 500 samples.

From a general examination of this table it appears that

1. There is not a great deal of difference in the performance of any of the variance estimators. They generally provide underestimates of the $MSE(\hat{R})$.
2. As would be expected, the error in the estimates tends to decrease as the sample size increases.

It might be noted that sampling errors are available for these entries that permit comparisons of a column (1) entry with a column (2) entry, within a row. Comparisons of entries within a row that are in the same column designations, (1) or (2), cannot be made because they are based on the same 500 samples and we did not compute the required covariances. An indication of their magnitudes is given in table C.

The stability of the variance estimators is examined in table C where a comparison of the mean square errors of the mean square error estimators is given. We observe that

1. For all the estimators there is a great deal of variability in relative performance over the different populations, with the population $y + 20$ causing particular difficulty for the bootstrap estimators.
2. With the exception of population $y + 20$, there does not appear to be much difference in the various estimators for $n = 10$ and $n = 20$. For $n = 50$, the bootstrap estimators generally seem less stable than the Taylor estimators.

Table C. Mean square errors of mean square error estimates by population and estimated parameters

[See text for explanation of symbols]

n and population	$\hat{V}_{\bar{Y}}(\hat{R})$		$\hat{V}_{\bar{Y}}(\hat{R})$		$\hat{V}_{BWR}(\hat{R})$		$\hat{V}_{BWO}(\hat{R})$		$M\hat{S}E_{BWR}(\hat{R})$		$M\hat{S}E_{BWO}(\hat{R})$	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
<i>n</i> = 10												
Basic.....	2.5(-5)	2.5(-5)	131	117	103	101	106	103				
<i>y</i> + 20.....	3.7(-3)	3.0(-3)	45	55	148	158	161	173				
<i>y</i> ^{1.5}	1.1(-1)	6.9(-2)	119	119	95	89	98	95				
<i>y</i> ^{1.5} + 20.....	3.5(-2)	4.6(-2)	128	132	100	92	101	96				
<i>y</i> + 1.5 <i>x</i> - 0.1 <i>x</i> ²	2.2(-4)	2.0(-4)	116	100	105	112	114	119				
Mean.....			108	107	110	110	116	117				
<i>n</i> = 20												
Basic.....	2.7(-6)	3.0(-6)	108	112	113	110	115	122				
<i>y</i> + 20.....	2.4(-4)	2.4(-4)	60	55	122	150	128	161				
<i>y</i> ^{1.5}	9.1(-3)	10.7(-3)	112	104	101	101	103	98				
<i>y</i> ^{1.5} + 20.....	5.7(-3)	6.0(-3)	116	111	97	107	98	109				
<i>y</i> + 1.5 <i>x</i> - 0.1 <i>x</i> ²	2.5(-5)	2.3(-5)	111	120	103	112	105	116				
Mean.....			101	100	107	116	110	121				
<i>n</i> = 50												
Basic.....	5.0(-8)	5.0(-8)	100	100	120	122	119	122				
<i>y</i> + 20.....	3.5(-6)	3.7(-6)	66	59	172	158	180	170				
<i>y</i> ^{1.5}	1.7(-4)	1.7(-4)	111	104	117	111	119	112				
<i>y</i> ^{1.5} + 20.....	1.0(-4)	1.0(-4)	107	105	107	114	109	115				
<i>y</i> + 1.5 <i>x</i> - 0.1 <i>x</i> ²	4.0(-7)	3.6(-7)	110	117	148	139	145	139				
Mean.....			99	97	133	129	134	132				

NOTES: \bar{V} values are based on 500 samples; BWR and BWO values are based on 100 bootstrap samples; $\hat{V}_{\bar{Y}}(\hat{R})$ taken as 100 percent.

On any line in this table, the entries in column (1) are based on 1 set of 500 samples; the entries in column (2) are based on an independent set of 500 samples.

The numbers in parentheses are the powers of 10 by which the first number is to be multiplied. Thus the first entry of 2.5(-5) is 0.000025.

The estimation of bias and an adjusted ratio estimator

With a single sample and the estimate \hat{R} , there is no way of estimating the ratio bias in \hat{R} , that is, $E(\hat{R}) - R$, without resorting to some pseudoreplication technique, such as the jack-knife. Efron³ (p. 33) and Efron and Gong² (p. 41) argue that the bootstrap provides an effective way of estimating bias. For a single sample, which becomes the population for bootstrap sampling, we have the ratio estimate \hat{R} . For each bootstrap sample we have either \hat{R}_{BWO}^* or \hat{R}_{BWR}^* , depending upon which bootstrap procedure we have chosen. Thus $\hat{R}_B^* - \hat{R}$ is an estimate of bias. Because we generally have drawn 100 bootstrap samples, we use $\bar{\hat{R}}_B^* - \hat{R}$ as an estimate of bias. Finally, the average of this quantity over the 500 samples drawn from the original population, $\bar{\hat{R}}_B^* - \hat{R}$, provides a still better estimate of bias. This quantity is shown in table D for our five populations and three sample sizes.

The general tendencies exhibited in this table are as follows:

1. With one exception, the sign of the estimated bias agrees with the actual sign.
2. Both the BWO and BWR procedures appear to give quite reasonable estimates of the actual bias, although the BWR scheme tends to provide overestimates for $n = 10$.

Although not shown in the table, these estimates of bias have small sampling errors. This occurs because \hat{R} and $\bar{\hat{R}}_B^*$ are obtained from a single sample and the correlation, over 500 samples, between \hat{R} and $\bar{\hat{R}}_B^*$ is of the order of 0.995. Furthermore, their estimated variances are quite close together. For example, for population $y + 20$ and $n = 10$

$$\bar{\hat{R}} = 3.3634$$

Table D. Estimation of bias of \hat{R} using the bootstrap

[See text for explanation of symbols]

Population	n = 10		n = 20		n = 50	
	$\bar{\hat{R}}_{BWO}^*$	$\bar{\hat{R}}_{BWR}^*$	$\bar{\hat{R}}_{BWO}^*$	$\bar{\hat{R}}_{BWR}^*$	$\bar{\hat{R}}_{BWO}^*$	$\bar{\hat{R}}_{BWR}^*$
Basic						
Bias	¹ -0.0002		² -0.0007		³ -0.0004	
Bootstrap estimate	-0.0007	-0.0009	0.0024	-0.0000	-0.0002	-0.0001
y + 20						
Bias	³ 0.0389		³ 0.0205		³ 0.0025	
Bootstrap estimate	0.0343	0.0395	0.0155	0.0172	0.0041	0.0043
y^{1.5}						
Bias	³ -0.0203		³ -0.0141		³ -0.0049	
Bootstrap estimate	-0.0272	-0.0355	-0.0129	-0.0121	-0.0046	-0.0043
y^{1.5} + 20						
Bias	³ 0.0103		³ 0.0092		³ 0.0017	
Bootstrap estimate	0.0071	0.0170	0.0045	0.0001	0.0011	0.0014
y + 1.5x - 0.1x²						
Bias	³ 0.0133		³ 0.0065		³ 0.0013	
Bootstrap estimate	0.0126	0.0140	0.0063	0.0055	0.0017	0.0015
Average absolute bias	0.0166		0.0102		0.0022	
Average estimate of absolute bias	0.0139	0.0214	0.0083	0.0072	0.0023	0.0023

¹Based on 15,000 samples.

²Based on 10,000 samples.

³Based on 5,000 samples.

NOTE: All others are based on 500 samples.

Table E. Bias of the adjusted ratio estimator

[See text for explanation of symbols]

Population	n = 10		n = 20		n = 50	
	BWO	BWR	BWO	BWR	BWO	BWR
Basic						
Actual bias of \hat{R}		-0.0002		-0.0007		-0.0004
Bias of \hat{R}_A^*	-0.0020	-0.0044	0.0097	-0.0159	-0.0007	-0.0097
$y + 20$						
Actual bias of \hat{R}		0.0389		0.0205		0.0025
Bias of \hat{R}_A^*	0.0124	-0.0046	0.0170	-0.0139	-0.0025	0.0013
$y^{1.5}$						
Actual bias of \hat{R}		-0.0203		-0.0141		-0.0049
Bias of \hat{R}_A^*	-0.0406	0.0028	0.0081	-0.0200	-0.0023	-0.0038
$y^{1.5} + 20$						
Actual bias of \hat{R}		0.0103		0.0092		0.0017
Bias of \hat{R}_A^*	-0.0026	-0.0436	-0.0106	0.0115	0.0052	-0.0045
$y + 1.5x - 0.1x^2$						
Actual bias of \hat{R}		0.0133		0.0065		0.0013
Bias of \hat{R}_A^*	0.0031	0.0081	0.0036	0.0025	-0.0034	0.0008
Average absolute bias.....		0.0166		0.0102		0.0022
Average estimate of absolute bias.....	0.0121	0.0127	0.0098	0.0128	0.0028	0.0040

NOTE: The actual bias values are based on 5,000 samples; the values of the bias of \hat{R}_A^* are based on 500 samples.

$$\bar{\bar{R}}_{BWO}^* = 3.3977$$

$$\text{Bias} = 0.0343$$

$$\hat{V}(\hat{R}) = 0.0910$$

$$\hat{V}(\bar{\bar{R}}_{BWO}^*) = 0.0961$$

$$\hat{\rho}_{\bar{R}, \bar{R}^*} = 0.9933$$

Therefore,

$$\hat{V}(\text{bias}) = \frac{0.0910}{500} + \frac{0.0961}{500}$$

$$- 2 \times 0.9933 \sqrt{\frac{0.0910}{500}} \sqrt{\frac{0.0961}{500}}$$

$$\sqrt{\hat{V}(\text{bias})} = 0.0016$$

From table D it appears that, at least approximately,

$$\bar{\bar{R}}_B^* - \bar{R} \doteq \bar{R} - R$$

This suggests that

$$\hat{R} - (\bar{\bar{R}}_B^* - \hat{R}) = 2\hat{R} - \bar{\bar{R}}^*$$

might be an “improved” estimator of R . This estimator will be denoted by $\hat{R}_{BWR,A}^*$ or $\hat{R}_{BWO,A}^*$, the subscript A denoting “adjusted.” The estimated bias of these two adjusted estimators for all five populations and three sample sizes is given in table E. These data do not seem to indicate that \hat{R}_A^* has smaller bias than \hat{R} because roughly half the cases show increased bias and half show decreased bias. However, the sampling errors of these estimates are of such magnitude that none of the differences between the estimates and the corresponding bias is significant. (The standard error of an estimate can be approximated by taking the appropriate $\hat{V}(\hat{R})$ from table A, dividing by 500, and taking the square root.)

Coverage of confidence intervals

For each sample used in this investigation we computed the following t -statistics:

1. BWR:

$$\begin{aligned}
 t_{T\bar{X}} &= \frac{\hat{R} - R}{\sqrt{\hat{V}_{T\bar{X}}(\hat{R})}} \\
 t_{T\bar{X}} &= \frac{\hat{R} - R}{\sqrt{\hat{V}_{T\bar{X}}(\hat{R})}} \\
 t_{BWR} &= \frac{\hat{R} - R}{\sqrt{\hat{V}_{BWR}(\hat{R})}} \\
 t_{MSE_{BWR}} &= \frac{\hat{R} - R}{\sqrt{MSE_{BWR}(\hat{R})}} \\
 t_{T\bar{X},A} &= \frac{2\hat{R} - \bar{R}_{BWR}^* - R}{\sqrt{\hat{V}_{T\bar{X}}(\hat{R})}} \\
 t_{T\bar{X},A} &= \frac{2\hat{R} - \bar{R}_{BWR}^* - R}{\sqrt{\hat{V}_{T\bar{X}}(\hat{R})}} \\
 t_{BWR,A} &= \frac{2\hat{R} - \bar{R}_{BWR}^* - R}{\sqrt{\hat{V}_{BWR}(\hat{R})}} \\
 t_{MSE_{BWR},A} &= \frac{2\hat{R} - \bar{R}_{BWR}^* - R}{\sqrt{MSE_{BWR}(\hat{R})}}
 \end{aligned} \tag{14}$$

2. BWO:

$$\begin{aligned}
 t_{T\bar{X}} &= \frac{2\hat{R} - \bar{R}_{BWO}^* - R}{\sqrt{\hat{V}_{T\bar{X}}(\hat{R})}} \\
 t_{T\bar{X}} &= \frac{2\hat{R} - \bar{R}_{BWO}^* - R}{\sqrt{\hat{V}_{T\bar{X}}(\hat{R})}} \\
 t_{BWO} &= \frac{2\hat{R} - \bar{R}_{BWO}^* - R}{\sqrt{\hat{V}_{BWO}(\hat{R})}} \\
 t_{MSE_{BWO}} &= \frac{2\hat{R} - \bar{R}_{BWO}^* - R}{\sqrt{MSE_{BWO}(\hat{R})}}
 \end{aligned} \tag{16}$$

Five hundred independent samples were drawn for each population, for each sample size, and for BWR and BWO. Table F presents the percent of the 500 samples for which the two-tailed t -statistic equaled or exceeded the tabular values for $\alpha = 0.05$ and 0.10 . Nine degrees of freedom were used for $n = 10$, 19 degrees of freedom for $n = 20$, and an infinite number of degrees of freedom for $n = 50$.

There are two important factors to keep in mind when looking at the entries in table F:

1. The numerator for any t in an A line is $\hat{R} - R$; the numerator for any t in a B line is $(2\hat{R} - \bar{R}_{BWR}^*) - R$; and the numerator for any t in a C line is $(2\hat{R} - \bar{R}_{BWO}^*) - R$. Actually, this difference in numerators has little effect because table E shows that there is very little difference among the estimators \hat{R} , $\bar{R}_{BWR,A}^*$, and $\bar{R}_{BWO,A}^*$.
2. The entries are two-tailed. We have not shown the one-tailed values, but they differ markedly from what would be expected under symmetry. For the first four populations, the percent of t 's that are smaller than $-t_\alpha$ is much larger than the percent that exceed $+t_\alpha$. For the fifth population $y + 1.5x - 0.1x^2$, the reverse is true.

The only tendency that seems to stand out in table F is that the entries generally exceed the nominal value. This means that confidence intervals will cover the parameter somewhat less frequently than would be expected from the stated confidence coefficient.

To bring out the more important aspects of table F, we have computed a number of summary tables, always keeping the numerators and denominators for each t -statistic separate and, of course, always keeping the two α values separate. In table G the comparison for variance estimators is presented. There are several points to note concerning this table:

1. There is very little difference among the four variance estimators.
2. There is a slight tendency for the $\bar{R}_{BWO,A}^*$ results to be larger than the other two.

In table H, comparison of t values among populations is given. The most striking feature of this table is the extent to which the t values for population $y^{1.5}$ exceed the nominal values. Again, we observe, as in table G, that the $\bar{R}_{BWO,A}^*$ results frequently are larger than the others.

Table J presents the comparisons for sample size. There seem to be no marked differences for the three sample sizes.

Table F. Percent of 500 samples for which the 2-tailed t-statistic equals or exceeds the tabular value

[See text for explanation of symbols]

Population	n	Method ¹	$\alpha = 0.05$				$\alpha = 0.10$			
			$\hat{V}_{T\bar{x}(\hat{R})}$	$\hat{V}_{T\bar{x}(\hat{R})}$	$\hat{V}_B(\hat{R})$	$M\hat{S}E_B(\hat{R})$	$\hat{V}_{T\bar{x}(\hat{R})}$	$\hat{V}_{T\bar{x}(\hat{R})}$	$\hat{V}_B(\hat{R})$	$M\hat{S}E_B(\hat{R})$
Basic.....	10	A	6.2	6.4	6.0	6.0	10.2	10.6	9.4	9.2
		B	5.6	6.4	5.8	5.8	10.6	11.0	9.8	9.8
		C	5.8	5.4	5.6	5.6	8.8	9.6	9.4	9.4
	20	A	4.2	4.8	4.6	4.6	9.8	9.6	10.6	10.2
		B	4.2	4.8	4.8	4.6	10.6	10.6	10.8	10.8
		C	7.3	8.0	7.4	7.3	13.1	12.8	14.5	14.3
	50	A	5.2	6.4	6.4	6.4	13.2	12.6	13.2	13.0
		B	6.4	7.2	7.0	6.8	14.0	14.2	14.0	13.8
		C	8.8	9.8	8.6	8.4	12.8	12.8	13.0	12.8
$y + 20$	10	A	7.4	6.8	7.0	7.0	12.6	12.2	11.6	11.4
		B	8.2	6.4	7.4	7.2	13.4	12.2	12.4	11.8
		C	4.4	4.6	4.0	4.0	7.6	8.8	9.0	8.8
	20	A	5.6	6.8	6.0	5.8	11.6	10.6	11.2	11.2
		B	5.6	6.6	6.0	5.6	12.4	10.8	13.6	12.4
		C	6.5	7.1	4.5	4.6	13.7	14.1	11.6	10.6
	50	A	4.2	4.6	4.6	4.4	9.4	8.8	8.4	8.2
		B	4.6	4.2	4.2	4.2	8.4	8.8	8.2	8.0
		C	4.8	5.2	5.8	5.6	10.6	10.0	10.0	10.0
$y^{1/2}$	10	A	7.6	8.8	7.0	7.0	12.2	13.2	11.8	11.6
		B	7.2	9.0	7.2	7.0	13.6	12.8	12.0	11.8
		C	6.4	7.6	8.0	8.0	13.4	13.0	14.4	14.0
	20	A	6.8	8.4	7.0	7.0	12.8	12.8	13.4	13.2
		B	6.0	8.4	6.8	6.6	13.8	13.4	14.0	13.4
		C	6.5	7.1	6.3	6.3	13.7	14.1	15.1	14.7
	50	A	8.6	8.8	8.4	8.2	13.8	13.2	14.8	14.6
		B	7.8	8.8	8.2	8.0	14.1	13.6	15.2	15.2
		C	8.2	8.2	9.2	8.8	12.6	12.6	13.4	13.2
$y^{1/2} + 20$	10	A	5.8	6.6	5.0	4.6	11.4	11.4	11.6	11.6
		B	5.2	6.8	5.4	5.4	12.2	11.8	11.8	11.2
		C	6.0	6.6	6.6	6.6	10.8	12.2	11.6	11.0
	20	A	6.0	6.0	6.4	6.4	11.0	11.2	11.4	11.4
		B	6.0	6.4	7.0	7.0	10.2	11.4	11.2	11.0
		C	5.5	6.4	6.7	6.1	11.7	12.3	12.3	12.0
	50	A	6.8	6.6	6.4	6.4	10.6	11.0	10.8	10.4
		B	6.6	6.8	6.4	6.4	10.6	10.6	12.2	11.8
		C	8.0	8.4	8.6	8.6	12.0	13.0	12.6	12.6
$y + 1.5x - 0.1x^2$	10	A	5.2	7.4	5.4	5.4	14.0	13.4	13.2	13.0
		B	5.4	6.6	5.8	5.8	13.6	12.6	13.2	12.6
		C	7.0	8.6	7.4	7.2	12.4	14.6	14.4	14.0
	20	A	8.2	9.4	9.8	8.8	14.8	14.4	14.4	14.4
		B	8.4	9.4	8.6	8.4	14.4	14.6	14.8	14.6
		C	8.0	9.4	9.3	9.1	15.2	15.6	16.1	15.6
	50	A	3.4	4.0	3.6	3.6	6.8	7.0	8.0	7.8
		B	3.4	3.8	3.6	3.6	7.2	7.4	7.8	7.8
		C	5.4	5.2	5.8	5.6	9.8	8.6	10.2	9.8

¹A: BWR and \hat{R} , formulas (14). B: BWR and $\hat{R}_{BWR,A}^* = 2\hat{R} - \hat{R}_{BWR}^*$, formulas (15). C: BWO and $\hat{R}_{BWO,A}^* = 2\hat{R} - \hat{R}_{BWO}^*$, formulas (16).

NOTE: For a particular population and sample size, all entries on lines A and B are based on the same set of 500 samples; entries on any other line in the table, including those in line C for the same population and sample size, are based on independent sets of 500 samples.

Table G. Comparison of effects of variance estimators from table F
 [See text for explanation of symbols]

Variance estimator	Ratio estimator	$\alpha = 0.05$	$\alpha = 0.10$
$\hat{V}_{\overline{YX}}(\hat{R})$	\hat{R}_{BWR}	6.1	11.6
	$\hat{R}_{BWR,A}^*$	6.0	11.9
	$\hat{R}_{BWO,A}^*$	6.6	11.9
$\hat{V}_{\overline{YX}}(\hat{R})$	\hat{R}_{BWR}	6.8	11.5
	$\hat{R}_{BWR,A}^*$	6.8	11.7
	$\hat{R}_{BWO,A}^*$	7.2	12.3
$\hat{V}_B(\hat{R})$	\hat{R}_{BWR}	6.2	11.6
	$\hat{R}_{BWR,A}^*$	6.3	12.1
	$\hat{R}_{BWO,A}^*$	6.9	12.5
$MSE_B(\hat{R})$	\hat{R}_{BWR}	6.1	11.4
	$\hat{R}_{BWR,A}^*$	6.2	11.7
	$\hat{R}_{BWO,A}^*$	6.8	12.2

Table H. Comparison of effects of different populations from table F
 [See text for explanation of symbols]

Population	Ratio estimator	$\alpha = 0.05$	$\alpha = 0.10$
Basic	\hat{R}_{BWR}	5.6	11.0
	$\hat{R}_{BWR,A}^*$	5.5	11.7
	$\hat{R}_{BWO,A}^*$	7.3	11.9
$y + 20$	\hat{R}_{BWR}	5.8	10.6
	$\hat{R}_{BWR,A}^*$	5.8	11.0
	$\hat{R}_{BWO,A}^*$	5.1	10.4
$y^{1.5}$	\hat{R}_{BWR}	7.8	13.1
	$\hat{R}_{BWR,A}^*$	7.6	13.6
	$\hat{R}_{BWO,A}^*$	7.6	13.7
$y^{1.5} + 20$	\hat{R}_{BWR}	6.1	11.2
	$\hat{R}_{BWR,A}^*$	6.3	11.3
	$\hat{R}_{BWO,A}^*$	7.0	12.0
$y + 1.5x - 0.1x^2$	\hat{R}_{BWR}	6.2	11.8
	$\hat{R}_{BWR,A}^*$	6.1	11.7
	$\hat{R}_{BWO,A}^*$	7.3	13.0

Table J. Comparison of effects of different sample sizes from table F

[See text for explanation of symbols]

Sample size	Ratio estimator	$\alpha = 0.05$	$\alpha = 0.10$
10	\hat{R}_{BWR}	6.4	11.8
	$\hat{R}_{BWR,A}^*$	6.5	12.0
	$\hat{R}_{BWO,A}^*$	6.3	11.4
20	\hat{R}_{BWR}	6.6	12.0
	$\hat{R}_{BWR,A}^*$	6.6	12.4
	$\hat{R}_{BWO,A}^*$	7.0	13.6
50	\hat{R}_{BWR}	5.8	10.8
	$\hat{R}_{BWR,A}^*$	5.9	11.6
	$\hat{R}_{BWO,A}^*$	7.4	11.6

Nonparametric confidence intervals for R

Efron³ (section 10.4) has suggested using $\hat{C}\hat{D}\hat{F}(\hat{\theta}^*)$ to determine nonparametric confidence intervals for θ . We apply this approach here to determine nonparametric confidence intervals for R .

Consider a sample of n drawn without replacement from N where R is the sample ratio. Draw samples with or without replacement from this sample according to the procedures previously suggested in this report. For a single such sample, denote the ratio estimator as \hat{R}^* . In the present instance we have used 1,000 bootstrap samples, each giving rise to an \hat{R}_i^* , $i = 1, \dots, 1,000$. Order these values of \hat{R}^* as

$$\hat{R}_{(1)}^* \leq \hat{R}_{(2)}^* \leq \dots \leq \hat{R}_{(1,000)}^*$$

For 95 percent confidence intervals, take $\alpha = 0.025$ and find \hat{R}_{lower}^* (\hat{R}_L^*) such that 2.5 percent of the \hat{R}^* values are $\leq \hat{R}_L^*$ and \hat{R}_{upper}^* (\hat{R}_U^*) such that 2.5 percent of the \hat{R}^* values are $\geq \hat{R}_U^*$. \hat{R}_L^* and \hat{R}_U^* provide a 95-percent nonparametric confidence interval for R . Ninety percent confidence intervals also were determined in the same manner. These percentile confidence intervals were computed for each of 500 samples drawn from the original population.

Efron³ (section 10.7) also describes a "bias-corrected percentile method" that applies here if the fraction of \hat{R}^* values $\leq \hat{R} \neq 0.50$. In the present instance, this involves the following steps. (The rationale for these steps is given by Efron.) For a single sample drawn from the original sample and the resulting 1,000 values of \hat{R}^* :

1. Find the fraction of values of $\hat{R}^* \leq \hat{R}$ that is $\hat{C}\hat{D}\hat{F}^*(\hat{R})$.
2. Find the standard normal variable z_0 such that the fraction of values of $z \leq z_0$ is equal to $\hat{C}\hat{D}\hat{F}^*(\hat{R})$.
3. Take z_α to be 1.96 for 95 confidence intervals and similarly for other confidence coefficients, and determine $2z_0 - z_\alpha = z_1$ and $2z_0 + z_\alpha = z_2$.
4. Determine the areas under a standard normal distribution such that α_1 is the probability that z is less than or equal to z_1 and α_2 is the probability that z is greater than or equal to z_2 , denoted by $\alpha_1 = \Pr(z \leq z_1)$ and $\alpha_2 = \Pr(z \geq z_2)$, respectively.
5. Find the bias-corrected percentile intervals $\hat{R}_{L,C}^*$ such that of the $\alpha_1 \times 100$ percent, \hat{R}^* values are $\leq \hat{R}_{L,C}^*$ and $\hat{R}_{U,C}^*$ such that of the $\alpha_2 \times 100$ percent, \hat{R}^* values are $\geq \hat{R}_{U,C}^*$.

Efron³ characterizes this procedure by arguing that it amounts to finding a transformation which transforms the distribution of \hat{R}^* to that of a normal distribution, if such a trans-

formation exists. It is not necessary to find the form of this transformation. Here we have determined both 95 and 90 percent confidence intervals.

To provide a basis of comparison, $\hat{V}_B^*(\hat{R})$ has also been computed from the 1,000 \hat{R}_i^* values, thus leading to the standard confidence intervals $\hat{R} \pm t\sqrt{\hat{V}_B^*(\hat{R})}$. All three intervals have been determined for the two populations $y^{1.5}$ and $y + 1.5x - 0.1x^2$ and for the three sample sizes 10, 20, and 50. For each case we also have employed the BWR and BWO procedures. These two populations were chosen because they seem to cause difficulty in coverage of confidence intervals, as indicated in table F. This is especially true for $y^{1.5}$.

In table K the average widths of the three 95-percent confidence intervals, together with the percent of 500 intervals that cover the population value R , are presented. These are given for the two populations, the three sample sizes, and for the BWR and BWO procedures. From this table we observe:

1. There is very little difference between the BWR and BWO procedures for width or coverage of the intervals.
2. There is very little difference between the percentile intervals and the bias-corrected percentile intervals.
3. All coverage percents are smaller than the nominal value, 95 percent.
4. For samples of size 10, the widths of the intervals for the two percentile methods are noticeably smaller than those of the ordinary method, $2t\sqrt{\hat{V}_B^*(\hat{R})}$. At the same time, however, the coverage of these intervals is smaller than that of the ordinary intervals. These differences still exist for samples of size 20, although they are less pronounced. They have completely disappeared for samples of size 50.
5. There is no marked improvement in the coverages as n increases from 10 to 50.

An analogous table for 90 percent confidence intervals yielded results consistent with these conclusions stated and, therefore, has not been included in this report.

Table K presents the major aspects of the differences among the three ways of determining confidence intervals. One other aspect, however, is worthy of brief consideration. In the preceding section it was observed that for the symmetric intervals considered there, the one-tailed coverages would be seriously in error. One might hope that the nonparametric intervals of this section would improve this situation. Although the intervals arising from these procedures are not symmetric, at least for small n , they do not produce more nearly equal tails. As an example, consider population $y + 1.5x - 0.1x^2$, $n = 10$, and

Table K. Coverages and average width of nonparametric 95 percent confidence intervals

[See text for explanation of symbols]

Population	n	Method	$2t\sqrt{\hat{V}_B^*(\hat{R})}$	$\hat{R}_U^* - \hat{R}_L^*$	$\hat{R}_{U,c}^* - \hat{R}_{L,c}^*$
$y + 1.5x - 0.1x^2$	10	BWR	0.64(94.8)	0.55(88.0)	0.54(87.8)
		BWO	0.62(92.2)	0.53(84.8)	0.53(85.8)
	20	BWR	0.41(93.2)	0.38(90.0)	0.38(90.4)
		BWO	0.41(93.2)	0.38(90.0)	0.38(89.4)
	50	BWR	0.20(93.4)	0.20(93.0)	0.20(92.8)
		BWO	0.20(89.4)	0.20(89.2)	0.20(89.0)
$y^{1.5}$	10	BWR	2.20(93.4)	1.88(90.6)	1.89(90.8)
		BWO	2.08(92.4)	1.78(89.2)	1.79(89.0)
	20	BWR	1.44(93.6)	1.34(92.2)	1.34(92.4)
		BWO	1.38(93.4)	1.29(90.6)	1.29(91.4)
	50	BWR	0.69(92.2)	0.69(91.6)	0.69(91.8)
		BWO	0.70(93.6)	0.69(93.0)	0.69(93.4)

NOTES: The entries on each line of the table are based on 500 samples from the population. Entries on different lines are based on independent sets of 500 samples.

The entries in parentheses are the percents of 500 confidence intervals that include R .

Each interval is based on 1,000 bootstrap samples; coverages and average widths are based on 500 samples from the population.

Table L. Location of R with respect to specified percentile intervals of \hat{R}^* with BWR

[See text for explanation of symbols]

Population	n	Percent of 500 samples for which—					
		$R \leq \hat{R}_{10.0}^*$	$\hat{R}_{10.0}^* < R \leq \hat{R}_{25.0}^*$	$\hat{R}_{25.0}^* < R \leq \hat{R}_{50.0}^*$	$\hat{R}_{50.0}^* < R \leq \hat{R}_{75.0}^*$	$\hat{R}_{75.0}^* < R \leq \hat{R}_{90.0}^*$	$\hat{R}_{90.0}^* < R$
$y^{1.5}$	10	7.0	13.0	23.6	24.0	14.8	17.8
	20	6.6	14.2	25.0	21.4	15.2	17.6
	50	5.6	15.2	25.4	19.4	14.8	19.6
$y + 1.5x - 0.1x^2$	10	20.6	14.6	22.2	25.8	10.2	6.6
	20	19.8	12.2	23.6	25.0	13.0	6.4
	50	12.6	13.6	26.0	23.0	17.8	7.0

NOTE: Based on 1,000 BWR samples.

BWR. For each sample from the population, we determined $\hat{R}_L^* - \hat{R}$ and $\hat{R}_U^* - \hat{R}$; also, $\hat{R}_{L,c}^* - \hat{R}$ and $\hat{R}_{U,c}^* - \hat{R}$. The values of these half-widths, averaged over 500 samples, are as follows:

Interval	Percentile
$\hat{R}_L^* - \hat{R}$	-0.24
$\hat{R}_U^* - \hat{R}$	0.30
$\hat{R}_{L,c}^* - \hat{R}$	-0.25
$\hat{R}_{U,c}^* - \hat{R}$	0.29

These compare with the symmetric intervals $\pm t\sqrt{\hat{V}_B^*(\hat{R})}$ of -0.32 and +0.32. For the percentile intervals, R was below \hat{R}_L^* in 10.2 percent of the samples, above \hat{R}_U^* in 1.8 percent of the samples, and between \hat{R}_L^* and \hat{R}_U^* in 88 percent of the samples. The corresponding values for the percentile-corrected intervals were 10.2, 2.0, and 87.8 percent. The same situation exists for the population $y^{1.5}$, except that the right tail of the distribution is the heavy tail instead of the left tail. This behavior does not improve with larger sample sizes.

Further information on the behavior of the percentiles of the distribution of \hat{R}_i^* is given in the following analysis of our basic data. For each sample drawn from a population we have

the ordered values of \hat{R}_i^* as indicated in equation (17). These were used to determine selected percentiles of the empirical distribution: $\hat{R}_{2.5}^*$, $\hat{R}_{5.0}^*$, $\hat{R}_{10.0}^*$, $\hat{R}_{25.0}^*$, $\hat{R}_{50.0}^*$, $\hat{R}_{90.0}^*$, $\hat{R}_{95.0}^*$, and $\hat{R}_{97.5}^*$. $\hat{R}_{2.5}^*$ and $\hat{R}_{97.5}^*$ determined the 95-percent nonparametric confidence intervals discussed earlier in this section. For each of the 500 samples, the location of R with respect to $\hat{R}_{10.0}^*$, $\hat{R}_{25.0}^*$, $\hat{R}_{50.0}^*$, $\hat{R}_{75.0}^*$, and $\hat{R}_{90.0}^*$ was observed. The percents of the 500 samples for which R fell in the stated intervals are given in table L. These values refer to BWR. The entire process was repeated for BWO, but because the results do not differ markedly from those given here they are not included.

From this table we observe the following:

1. R falls in one tail too high a percent of the time and in the other tail too low a percent of the time. The heavy tails are reversed in the two populations.
2. This situation improves somewhat with increasing sample size for population $y + 1.5x - 0.1x^2$, but appears to worsen with increasing sample size for population $y^{1.5}$.

Rao and Wu⁸ suggest that a second phase bootstrap procedure on t values might provide improved confidence intervals, but this proposal did not come to our attention soon enough to be included in this study.

Two-stage sampling

In cluster sampling, where one first selects a sample of clusters and then draws a sample of elements from each of the selected clusters, the variability of an estimate of a population parameter depends upon the variability between clusters and the variability within clusters. Thus, Cochran⁹ (chapter 10) presents the theory for the sampling and subsampling of equal-size clusters. For example, if the population consists of N clusters, each with M elements, and a simple random sample of n clusters is selected, with m elements drawn at random from each, we have for the sample mean

$$V(\bar{y}) = \frac{N-n}{N} \frac{1}{n} \frac{\sum_{i=1}^N (\bar{Y}_i - \bar{Y})^2}{N-1} + \frac{M-m}{M} \frac{1}{nm} \frac{\sum_{i=1}^N \sum_{j=1}^M (y_{ij} - \bar{Y}_i)^2}{N(M-1)} \quad (17)$$

and

$$\hat{V}(\bar{y}) = \frac{N-n}{N} \frac{1}{n} \frac{\sum_{i=1}^n (\bar{y}_i - \bar{y})^2}{n-1} + \frac{n}{N} \frac{M-m}{M} \frac{1}{nm} \frac{\sum_{i=1}^n \sum_{j=1}^m (y_{ij} - \bar{y}_i)^2}{n(m-1)} \quad (18)$$

To estimate $V(\bar{y})$ from a single sample it is necessary to estimate the first and second terms separately and add them together. This is true whether one uses a Taylor series approximation or a pseudoreplication technique, such as balanced half samples or the jackknife.¹¹

Various attempts have been made to devise a pseudoreplication technique where the variability between replicate estimates would account correctly for the between and within components without the necessity of estimating them separately. These attempts have been unsuccessful; see, for example, McCarthy¹¹ and Mellor¹² (p. 189). However, it is possible to use the bootstrap to accomplish this end. We will describe BWO and BWR schemes, but the BWO scheme is not of value except in very special circumstances.

If $N = k_1 n$ and $M = k_2 m$, then the estimated variance $\hat{V}(\bar{y})$ of a sample mean

$$\bar{y} = \frac{\sum_{i=1}^n \sum_{j=1}^m y_{ij}}{nm}$$

is equal to

$$\hat{V}(\bar{y}) = \left(1 - \frac{1}{k_1}\right) \frac{1}{n} \frac{\sum_{i=1}^n (\bar{y}_i - \bar{y})^2}{n-1} + \frac{1}{k_1} \left(1 - \frac{1}{k_2}\right) \frac{1}{nm} \frac{\sum_{i=1}^n \sum_{j=1}^m (y_{ij} - \bar{y}_i)^2}{n(m-1)} \quad (19)$$

Suppose that we now replicate each cluster k_1 times and draw a sample of n clusters without replacement from this artificial population. For the selected clusters, replicate each element $k_1 k_2$ times and draw, without replacement, a sample of $k_1 m$ elements. Then the variance of the bootstrap sample mean \bar{y}^* , given the original sample, is

$$\begin{aligned} V^*(\bar{y}^* | \text{original sample}) &= \hat{V}_{\text{BWO}}(\bar{y}) \\ &= \left(1 - \frac{1}{k_1}\right) \frac{1}{n} \frac{k_1 \sum_{i=1}^n (\bar{y}_i - \bar{y})^2}{k_1 n - 1} \\ &\quad + \left(1 - \frac{1}{k_2}\right) \frac{1}{nk_1 m} \\ &\quad \times \frac{k_1^2 k_2 \sum_{i=1}^n \sum_{j=1}^m (y_{ij} - \bar{y}_i)^2}{nk_1 (k_1 k_2 m - 1)} \\ &= \left(1 - \frac{1}{k_1}\right) \frac{1}{n} \frac{\sum_{i=1}^n (\bar{y}_i - \bar{y})^2}{n - 1/k_1} \end{aligned}$$

$$+ \left(1 - \frac{1}{k_2}\right) \frac{1}{k_1} \frac{1}{nm}$$

$$\times \frac{\sum_{i=1}^n \sum_{j=1}^m (y_{ij} - \bar{y}_i)^2}{n(m - 1/k_1 k_2)} \quad (20)$$

When we compare $\hat{V}(\bar{y})$ and $V(\bar{y}^* | \text{original sample})$, we see that the correction factors needed to make the two equal are different for the within and between components: $(n - 1/k_1)/(n - 1)$ for the between component and $(m - 1/k_1 k_2)/(m - 1)$ for the within component. Thus, the only situations in which this would be a practical approach are those in which these two factors are approximately equal or, better still, where they are both approximately equal to 1.

For BWR, suppose we draw n^* clusters with replacement from the n sample clusters and m^* elements with replacement from the m elements in each of the n^* clusters selected at the first stage. Then the variance of the sample mean is given by

$$V^*(\bar{y}^* | \text{original sample}) = \frac{1}{n^*} \frac{\sum_{i=1}^n (\bar{y}_i - \bar{y})^2}{n} + \frac{1}{n^* m^*}$$

$$\times \frac{\sum_{i=1}^n \sum_{j=1}^m (y_{ij} - \bar{y}_i)^2}{nm} \quad (21)$$

This will be equal to $\hat{V}(\bar{y})$, as given by equation (19), if we choose n^* and m^* so that

$$\left(1 - \frac{1}{k_1}\right) \frac{1}{n} \frac{1}{n-1} = \frac{1}{n^* n}$$

or

$$n^* = \frac{n-1}{1-1/k_1}$$

$$= \frac{k_1}{k_1-1} (n-1) \quad (22)$$

and

$$\frac{1}{k_1} \left(1 - \frac{1}{k_2}\right) \frac{1}{nm} \frac{1}{n(m-1)} = \frac{1}{n^* m^* nm}$$

or

$$m^* = \frac{k_1 k_2}{k_2 - 1} \frac{n}{n^*} (m - 1) \quad (23)$$

Thus, with this choice of n^* and m^* ,

$$\hat{V}^*(\bar{y}^* | \text{original sample}) = \hat{V}_{\text{BWR}}(\bar{y}) = \hat{V}(\bar{y})$$

and it is only necessary to draw repeated bootstrap samples from the original sample, make the estimate \bar{y}^* from each sample, and compute the variance among these estimates.

No attempt has been made to study this procedure when k_1 and k_2 are not integers or when n^* and m^* are not integers.

The approach of Rao and Wu,⁸ described earlier in this report, also can be extended to two-stage sampling.

To try out the BWR scheme on actual data, we developed a reasonable superpopulation model from which an actual finite population could be drawn. We started with

$$y_{ij} = \mu + c_i + e_{ij}$$

where c_i is the cluster effect and e_{ij} controls the variability within clusters. Let c_i have mean zero and variance σ_b^2 , and e_{ij} have mean zero and variance σ_w^2 , c_i and e_{ij} being independent. Then

$$V(y_{ij}) = \sigma_b^2 + \sigma_w^2$$

and the intraclass correlation coefficient between elements in the same cluster is

$$\rho_{y_{ij}, y_{ik}} = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2}$$

We took c_i to be χ^2 with nine degrees of freedom and e_{ij} to be normal with mean of zero and variance σ_w^2 . ρ was arbitrarily assigned the value 0.10 because this appears to be a value that might well arise in practice. This leads to $\sigma_w^2 = 162$. Finally, μ was taken to be 40 to insure that y_i is greater than zero.

Because we wanted to apply the bootstrap method to the ratio estimator, it was next necessary to generate an x_{ij} to be associated with each y_{ij} . We started with the customary model

$$y_{ij} = R x_{ij} + e'_{ij}$$

$$E(e'_{ij}) = 0$$

$$V(e'_{ij}) = \sigma_w'^2$$

with e'_{ij} normal and independent of c_i and e_{ij} thus leading to

$$x_{ij} = \frac{y_{ij}}{R} + \frac{e'_{ij}}{R}$$

$$V(x_{ij}) = \frac{1}{R^2} (\sigma_b^2 + \sigma_w^2 + \sigma_w'^2)$$

Taking $R = 2$ and $\sigma_w'^2 = 90$, we find

$$\rho_{xy} = 0.82$$

$$\rho_{x_{ij}, x_{ik}} = 0.067$$

Using this model, we generated two different populations to obtain information on different cluster sizes and sample

sizes. The characteristics of these populations are as follows:

Parameter	Population	
	A	B
N	15	10
M	12	18
k_1	3	2
k_2	6	6
n	5	5
m	2	3
n^*	6	8
m^*	3	3

These two populations are shown in figures 4 and 5. We also added 50 to each y value for each population, leaving the x values unchanged, thus leading to populations A + 50 and B + 50. Some of the characteristics of these four populations are given in table M.

Five hundred samples were drawn from each population and 100 bootstrap samples were selected from each sample. The following estimates of mean square error were computed for each sample:

$$\hat{V}_{TX}(\hat{R}) = \frac{1}{\bar{x}^2} \left[\frac{N-n}{N} \frac{1}{n} \frac{\sum_{i=1}^n (\bar{y}_i - \hat{R}\bar{x}_i)^2}{n-1} + \frac{n}{N} \frac{M-m}{M} \frac{1}{nm} \times \frac{\sum_{i=1}^n \sum_{j=1}^m (y_{ij} - \hat{R}x_{ij})^2}{n(m-1)} \right] \quad (24)$$

$$\hat{V}_{TX}(\hat{R}) = \frac{1}{\bar{x}^2} \left[\frac{N-n}{N} \frac{1}{n} \frac{\sum_{i=1}^n (\bar{y}_i - \hat{R}\bar{x}_i)^2}{n-1} + \frac{n}{N} \frac{M-m}{M} \frac{1}{nm} \times \frac{\sum_{i=1}^n \sum_{j=1}^m (y_{ij} - \hat{R}x_{ij})^2}{n(m-1)} \right] \quad (25)$$

$$\hat{V}_{BWR}^*(\hat{R}) = \frac{\sum_{i=1}^{100} (\hat{R}_i^* - \bar{\hat{R}}^*)^2}{99} \quad (26)$$

$$MSE_{BWR}^*(\hat{R}) = \hat{V}_{BWR}^*(\hat{R}) + (\bar{\hat{R}}^* - \hat{R})^2 \quad (27)$$

These values were averaged over the 500 samples, giving rise to $\bar{V}_{TX}(\hat{R})$, $\bar{V}_{TX}(\hat{R})$, $\bar{V}_{BWR}^*(\hat{R})$, and $MSE_{BWR}^*(\hat{R})$.

These values were compared with the $MSE(\hat{R})$ as computed from table M and expressed as the percent deviations presented in table N. Thus, for population A and $\bar{V}_{BWR}^*(\hat{R})$,

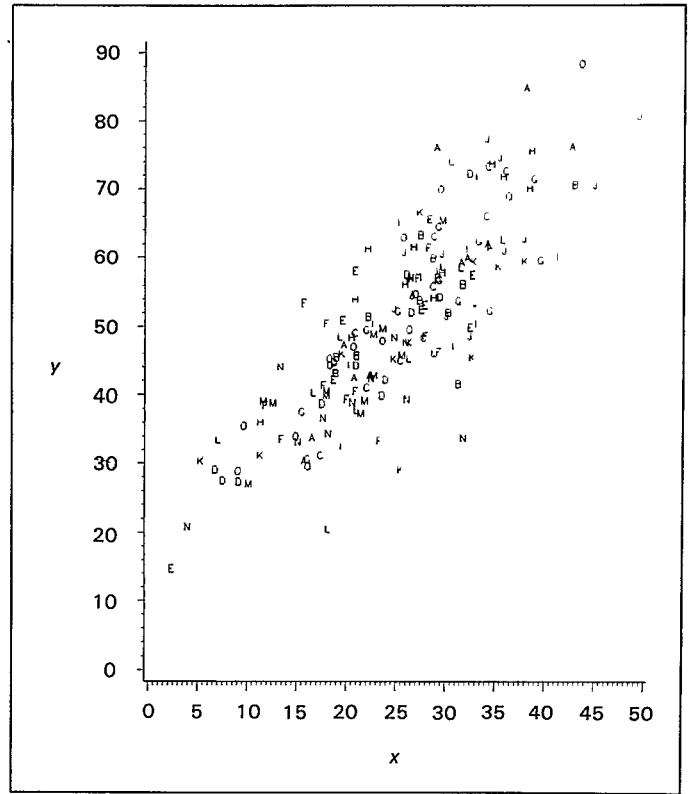


Figure 4. Scatter diagram of 15 clusters and 12 elements for population A (coordinates with the same letter are elements of the same cluster)

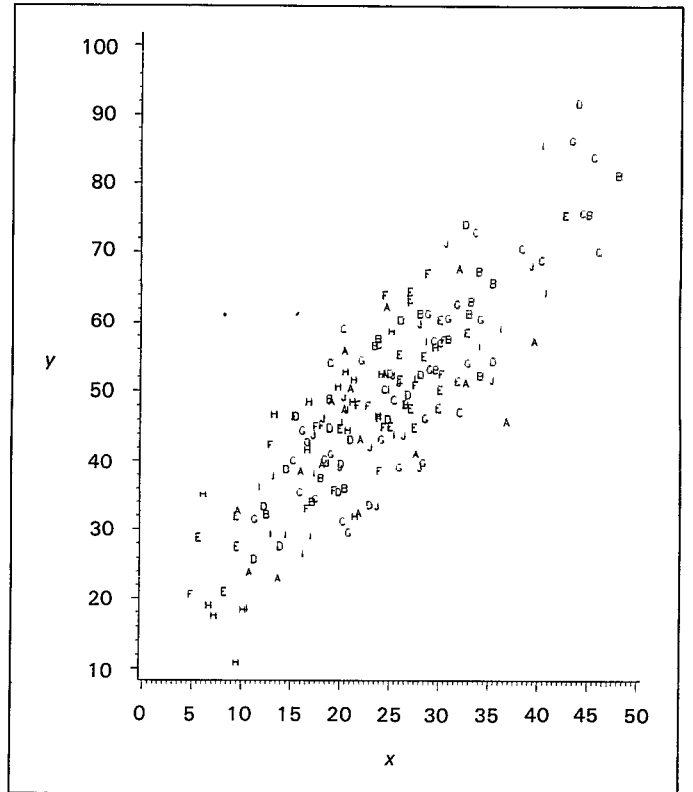


Figure 5. Scatter diagram of 10 clusters and 18 elements for population B (coordinates with the same letter are elements of the same cluster)

Table M. Populations used in the investigation of clustered populations and samples

[See text for explanation of symbols]

Population ¹	R	\bar{R}	Bias	$\hat{V}(\hat{R})$	$\frac{\hat{Bias}}{\sqrt{\hat{V}(\hat{R})}}$
					Percent
A	2.0113	2.0173	0.0060	0.01353	5.2
A + 50	3.9959	4.0283	0.0283	0.09984	9.0
B	2.0149	2.0228	0.0079	0.01056	7.7
B + 50	4.1062	4.1288	0.0226	0.07914	8.0

¹Population A: $N = 15$, $M = 12$, $n = 5$, and $m = 2$. Population B: $N = 10$, $M = 18$, $n = 5$, and $m = 3$.

NOTE: Based on 5,000 samples

Table N. Underpercent and overpercent estimates of $MSE(\hat{R})$ for clustered populations and samples

[See text for explanation of symbols]

Population ¹	$\bar{V}_{TX}(\hat{R})$	$\bar{V}_{TX}(\hat{R})$	$\bar{V}_{BWR}(\hat{R})$	$\overline{MSE}_{BWR}(\hat{R})$
A	3.3	- 0.4	12.1	14.3
A + 50	-2.2	-11.0	12.0	14.9
B	-7.1	-10.4	-3.2	-1.9
B + 50	-1.3	-5.6	4.9	7.9

¹Population A: $N = 15$, $M = 12$, $n = 5$, and $m = 2$. Population B: $N = 10$, $M = 18$, $n = 5$, and $m = 3$.

NOTE: Values of $MSE(\hat{R})$ are based on 5,000 samples; \hat{V} values are based on 500 samples; BWR values are based on 100 bootstrap samples.

$$\frac{\bar{V}_{BWR}^*(\hat{R}) - 0.01357}{0.01357} \times 100 \text{ percent} = 12.1 \text{ percent}$$

From this table we see that

1. The Taylor series estimators generally underestimate $MSE(\hat{R})$; the bootstrap estimators are overestimates.
2. Populations A and A + 50 cause some problem for the two bootstrap estimators, but they do well for populations B and B + 50.

If desired, sampling errors can be approximated from the entries in table M and the first column in table O.

The stability of the variance estimators is examined in table O. As was the case with \hat{R} for simple random sampling (see earlier discussion), the two bootstrap estimators are considerably less stable than the two Taylor series estimators. $\overline{MSE}_{BWR}(\hat{R})$ generally is the least stable of the four.

As discussed earlier, $\bar{R}_B^* - \hat{R}$ should provide an estimate of the bias of \hat{R} . This quantity, averaged over 500 samples ($\bar{R}_B^* - \hat{R}$), is presented in table P, where it is compared with the

actual bias as determined from 5,000 samples. The estimated bias agrees quite well with the actual bias in all cases.

For each of the 500 samples, for each of the populations, we computed the respective t -statistic from equations (14). In assessing the coverage of confidence intervals based on these t values, there occurs the problem of choosing the appropriate number of degrees of freedom. There seems to be no obvious answer to this concern. For populations A and A + 50, a variance estimate is based on a weighted sum of a between component (based on five clusters) and a within component (based on two elements per cluster). Therefore, it would seem reasonable that the number of degrees of freedom should be between five and nine (five for within plus four for between). We have assumed arbitrarily that the t values have seven degrees of freedom. Some empirical evidence that this choice is reasonable, in connection with a closely related problem, is presented by McCarthy.¹¹ For populations B and B + 50, the same line of reasoning leads to using a t value with 12 degrees of freedom.

Table Q gives the percent of the 500 samples for which the two-tailed t -statistic equals or exceeds the tabular value. These values suggest that quite reasonable confidence intervals for R can be obtained using any of the variance estimators.

Table O. Mean square errors of mean square error estimates for clustered populations and samples

[See text for explanation of symbols]

Population ¹	$\bar{V}_{TX}(\hat{R})$	$\bar{V}_{TX}(\hat{R})$	$\bar{V}_{BWR}(\hat{R})$	$\overline{MSE}_{BWR}(\hat{R})$
A	7.8(-5)	68	150	162
A + 50	5.7(-3)	51	184	203
B	2.2(-5)	69	121	125
B + 50	1.6(-3)	69	135	144

¹Population A: $N = 15$, $M = 12$, $n = 5$, and $m = 2$. Population B: $n = 10$, $M = 18$, $n = 5$, and $m = 3$.

NOTES: The numbers in parentheses are the powers of 10 by which the first number is to be multiplied. Thus the first entry of 7.8(-5) is 0.000078.

\bar{V} values are based on 500 samples; BWR values are based on 100 bootstrap samples; $\bar{V}_{TX}(\hat{R})$ taken as 100 percent.

Table P. Comparison of actual and estimated bias of \hat{R} by each population using the bootstrap for cluster sampling

Population ¹	Actual bias ²	Estimated bias ³
A	0.0060	0.0088
A + 50	0.0283	0.0326
B	0.0079	0.0057
B + 50	0.0226	0.0245

¹Population A: $N = 15$, $M = 12$, $n = 5$, and $m = 2$. Population B: $N = 10$, $M = 18$, $n = 5$, and $m = 3$.

²Actual bias based on 5,000 samples for each population.

³Estimated bias based on 500 samples for each population with 100 bootstrap samples for each sample.

Table Q. Percent of 500 samples for which 2-tailed *t*-statistic equals or exceeds the tabular value

[See text for explanation of symbols]

Population ¹	Denominator of <i>t</i>							
	$\hat{V}_{T\bar{x}}(\hat{R})$		$\hat{V}_{T\bar{x}}(\hat{R})$		$\hat{V}_{BWR}(\hat{R})$		$M\hat{S}E_{BWR}(\hat{R})$	
	5 percent	10 percent	5 percent	10 percent	5 percent	10 percent	5 percent	10 percent
A	4.0	10.6	4.2	10.6	4.4	9.3	4.2	9.0
A + 50.....	6.0	10.2	6.2	10.0	5.8	9.6	5.6	9.6
B	4.6	11.0	4.2	10.4	3.4	11.2	3.4	11.2
B + 50.....	5.4	9.8	5.0	10.4	5.2	9.4	5.2	9.2

¹Population A: $N = 15$, $M = 12$, $n = 5$, and $m = 2$. Population B: $N = 10$, $M = 18$, $n = 5$, and $m = 3$.

NOTE: 7 degrees of freedom for populations A and A + 50; 12 degrees of freedom for populations B and B + 50.

Summary

Two procedures are suggested for adapting Efron's bootstrap to finite population sampling. With the first procedure, sample elements are replicated to create an artificial population from which repeated samples are drawn without replacement. With the second procedure, sample elements are drawn with replacement from the original sample, the sample size being chosen so as to account for the finite population correction.

Using simulations based on five artificial populations and the ratio estimator, these two procedures are compared with each other and with two Taylor series variance estimators with regard to

1. The estimation of the mean square error of \hat{R} .
2. The stability of the variance estimators.
3. The coverage of confidence intervals obtained from the variance estimators.

In very general terms, there appear to be small differences among the variance estimators, except that the bootstrap estimators are somewhat less stable than the Taylor series estimators. The confidence intervals usually cover the population parameter a smaller fraction of times than would be expected from the nominal values.

The bootstrap procedures are employed in obtaining estimates of the ratio bias of \hat{R} and to produce nonparametric confidence intervals for R . This is not possible with an ordinary

single-sample Taylor series estimation of the variance approach. The estimates of bias are quite good, but the nonparametric confidence intervals do somewhat worse, at least for small samples, than the ordinary approach where coverage probabilities are concerned.

The BWR procedure is applied to the two-stage cluster sampling situation. It is possible to draw, with replacement, clusters from the sample clusters and elements from within these clusters in such a manner that the variance of the bootstrap means, in the linear case, is equal to the ordinary variance estimator. No separate estimation of the within- and between-cluster components is required. This procedure is evaluated by simulation with a number of artificial populations using the ratio estimator. It compares favorably with the ordinary Taylor series approximations in variance estimation and coverage of confidence intervals. It also produces a good estimate of the ratio bias in \hat{R} . The BWO procedure cannot be used in this situation unless very special conditions hold.

It is clear that the BWR schemes will generalize directly to stratified and stratified cluster sampling situations. BWO does not generalize to these except in very special circumstances. It also is possible to deal directly with medians by using the bootstrap.

Further work in adapting the bootstrap to sampling with unequal probabilities and without replacement from a finite population is planned.

References

- ¹B. Efron: Bootstrap methods: Another look at the jackknife. *Ann. Stat.* 7:1-26, 1979.
- ²B. Efron and G. Gong: A leisurely look at the bootstrap, the jackknife and cross-validation. *Am. Statistician* 37:36-48, 1983.
- ³B. Efron: The jackknife, the bootstrap, and other resampling plans. *Soc. Ind. Appl. Math.* Monograph No. 38, 1982.
- ⁴P. J. Bickel and D. A. Freedman: Some asymptotic theory for the bootstrap. *Ann. Stat.* 9:1196-1217, 1981.
- ⁵P. Diaconis and B. Efron: Computer intensive methods in statistics. *Sci. Am.* 248:116-130, 1983.
- ⁶P. J. Bickel and D. A. Freedman: Asymptotic Normality and the Bootstrap in Stratified Sampling. Unpublished manuscript. University of California at Berkeley, 1983.
- ⁷S. T. Gross: Median estimation in sample surveys. *Proceedings of the American Statistical Association, Section on Survey Research Methods*, 1980.
- ⁸J. N. K. Rao and C. F. J. Wu: Bootstrap Inference with Stratified Samples. Unpublished manuscript. Carleton University, Ottawa, and University of Wisconsin at Madison, 1983.
- ⁹W. G. Cochran: *Sampling Techniques*. New York. John Wiley & Sons, Inc., 1977.
- ¹⁰B. Efron: Censored data and the bootstrap. *JASA* 76:312-319, 1981.
- ¹¹P. J. McCarthy: Estimated Variance for the Combined Ratio Estimate: Stratified, Two-stage Samples Without Replacement. B. V. Sukhatme Memorial Lecture. Iowa State University, Apr. 16, 1982.
- ¹²R. W. Mellor: Subsample Replication Variance Estimator. Unpublished doctoral dissertation. Harvard University, 1973.

Vital and Health Statistics series descriptions

- SERIES 1. Programs and Collection Procedures**—Reports describing the general programs of the National Center for Health Statistics and its offices and divisions and the data collection methods used. They also include definitions and other material necessary for understanding the data.
- SERIES 2. Data Evaluation and Methods Research**—Studies of new statistical methodology including experimental tests of new survey methods, studies of vital statistics collection methods, new analytical techniques, objective evaluations of reliability of collected data, and contributions to statistical theory. Studies also include comparison of U.S. methodology with those of other countries.
- SERIES 3. Analytical and Epidemiological Studies**—Reports presenting analytical or interpretive studies based on vital and health statistics, carrying the analysis further than the expository types of reports in the other series.
- SERIES 4. Documents and Committee Reports**—Final reports of major committees concerned with vital and health statistics and documents such as recommended model vital registration laws and revised birth and death certificates.
- SERIES 5. Comparative International Vital and Health Statistics Reports**—Analytical and descriptive reports comparing U.S. vital and health statistics with those of other countries.
- SERIES 10. Data From the National Health Interview Survey**—Statistics on illness, accidental injuries, disability, use of hospital, medical, dental, and other services, and other health-related topics, all based on data collected in the continuing national household interview survey.
- SERIES 11. Data From the National Health Examination Survey and the National Health and Nutrition Examination Survey**—Data from direct examination, testing, and measurement of national samples of the civilian noninstitutionalized population provide the basis for (1) estimates of the medically defined prevalence of specific diseases in the United States and the distributions of the population with respect to physical, physiological, and psychological characteristics and (2) analysis of relationships among the various measurements without reference to an explicit finite universe of persons.
- SERIES 12. Data From the Institutionalized Population Surveys**—Discontinued in 1975. Reports from these surveys are included in Series 13.
- SERIES 13. Data on Health Resources Utilization**—Statistics on the utilization of health manpower and facilities providing long-term care, ambulatory care, hospital care, and family planning services.
- SERIES 14. Data on Health Resources: Manpower and Facilities**—Statistics on the numbers, geographic distribution, and characteristics of health resources including physicians, dentists, nurses, other health occupations, hospitals, nursing homes, and outpatient facilities.
- SERIES 15. Data From Special Surveys**—Statistics on health and health-related topics collected in special surveys that are not a part of the continuing data systems of the National Center for Health Statistics.
- SERIES 20. Data on Mortality**—Various statistics on mortality other than as included in regular annual or monthly reports. Special analyses by cause of death, age, and other demographic variables; geographic and time series analyses; and statistics on characteristics of deaths not available from the vital records based on sample surveys of those records.
- SERIES 21. Data on Natality, Marriage, and Divorce**—Various statistics on natality, marriage, and divorce other than as included in regular annual or monthly reports. Special analyses by demographic variables; geographic and time series analyses; studies of fertility; and statistics on characteristics of births not available from the vital records based on sample surveys of those records.
- SERIES 22. Data From the National Mortality and Natality Surveys**—Discontinued in 1975. Reports from these sample surveys based on vital records are included in Series 20 and 21, respectively.
- SERIES 23. Data From the National Survey of Family Growth**—Statistics on fertility, family formation and dissolution, family planning, and related maternal and infant health topics derived from a periodic survey of a nationwide probability sample of ever-married women 15-44 years of age.

For a list of titles of reports published in these series, write to

Scientific and Technical Information Branch
National Center for Health Statistics
Public Health Service
Hyattsville, Md. 20782

or call 301-436-NCHS

U S DEPARTMENT OF HEALTH AND
HUMAN SERVICES
Public Health Service
National Center for Health Statistics
3700 East-West Highway
Hyattsville, Maryland 20782

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE, \$300

THIRD CLASS MAIL
BULK RATE
POSTAGE & FEES PAID
PHS/NCHS
PERMIT No. G-281