Data Evaluation and
Methods Research

Series 2
Number 14

Replication

An Approach to the Analysis of Data From Complex Surveys

Development and evaluation of a replication technique for estimating variance.

DHEW Publication No. (PHS) 79-1269

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
Public Health Service
Office of Health Research, Statistics, and Technology
National Center for Health Statistics
Hyattsville, Maryland 20782

Vital and Health Statistics Series 2-No. 14
First issued as DHEW Publication No. (PHS) 66-1000
April 1966

NATIONAL CENTER FOR HEALTH STATISTICS

Library of Congress Catalog Card Number 66-60010

PREFACE

The theory of design of surveys has advanced greatly in the past three decades. One result is that many surveys now rest upon complex designs involving such factors as stratification and poststratification, multistage cluster sampling, controlled selection, and ratio, regression, or composite estimation. Another result is a growing concern and search for valid and efficient techniques for analysis of the output from such complex surveys.

A central difficulty is that most of the standard classical techniques for statistical analysis assume that observations are independent of one another and are the result of simple random sampling, often from a universe of normal or other known distribution—a situation that does not prevail in modern complex design. This report reviews several aspects of the problem and the limited literature on the topic. It offers a new method of balanced half-sample pseudoreplication as a solution to one phase of the problem.

The entire matter of how best to analyze data from complex surveys is nearly as broad as statistical theory itself. It encompasses not only the technical features of analysis, but also relationships among purpose and design of the survey, and the character of inferences which may be drawn about populations other than the finite universe which was sampled. This report treats only a very small sector of the subject, but, it is believed, introduces a scheme which may be widely useful. There is good reason to hope that the method, or possibly variations of it, may have utility beyond the somewhat narrow area with which it deals specifically.

The exploration and developments reported here are the outgrowth of discussions among a number of people, as is nearly always true when the subject is a pervasive one. But they are particularly the product of a study by Philip J. McCarthy of Cornell University under a contractual arrangement with the National Center for Health Statistics. Contributions of the Center were coordinated by Walt R. Simmons. Professor McCarthy wrote the report. Garrie J. Losee of the Center was responsible for initial work on half-sample replication, as employed in NCHS surveys, and prepared the appendix to this report.

CONTENTS

A key feature of statistical techniques necessary to the analysis of data from complex surveys is the method of calculating variance of the sample estimates. Earlier direct computational procedures are either inappropriate or much too difficult, even with high speed electronic computers, to cope with the elaborate stratification, multistage cluster sampling, and intricate estimation schemes found in many current sample surveys. A different approach is needed.

A number of statisticians have attempted solution through a variety of schemes which employ some form of replication or random grouping of observations. These efforts are recalled in this report, as a part of the background review of principal issues present in choice of analytic methods suitable to the complex survey.

Among the estimating schemes in recent use is a half-sample pseudoreplication technique adapted by the National Center for Health Statistics from an approach developed by the U.S. Bureau of the Census. This method is described in detail in the report. Typically, it involves subsampling a parent sample in such a way that 20-40 pseudoreplicated estimates of any specified statistic are produced, with the precision of the corresponding statistic from the parent sample being estimated from the variability among the replicated estimates.

One difficulty in using this method is that the 20-40 estimates are chosen from among the thousands or millions of possible replicates of the same character, and hence may yield an unstable estimate of the variability among the possible replicates. The report presents a system for controlled choice of a limited number of pseudoreplicates—often no more than 20-40 for a major national survey—such that for some classes of statistics the chosen small number of replicates has a variance algebraically identical with that of all possible replicates of the same character within the parent sample, and the same expected value as the variance of all possible replicates of the same character for all possible parent samples of the same design. Illustrations of the technique and guides for its use are included.

SYMBOLS

Data not available------------------------	---
Category not applicable------------------	...
Quantity zero----------------------------	-
Quantity more than 0 but less than 0.05-----	0.0
Figure does not meet standards of reliability or precision------------------	*

REPLICATION

AN APPROACH TO THE ANALYSIS OF DATA FROM COMPLEX SURVEYS

Philip J. McCarthy, Ph. D., *Cornell University*

INTRODUCTION

A considerable body of theory and practice has been developed relating to the design and analysis of sample surveys. This material is available in such books as Cochran (1963), Deming (1950), Hansen, Hurwitz, and Madow (1953), Kish (1965), Sukhatme (1954), and Yates (1960), and in numerous journal articles. Much of this theory and practice has the following characteristics: the sampled populations contain finite numbers of elements; no assumptions are made concerning the distributions of the pertinent variables in the population; major emphasis is placed on the estimation of simple population parameters such as percentages, means, and totals; and the samples are assumed to be "large" so that the sampling distributions of estimates can be approximated by normal distributions. Furthermore, it has frequently been appropriate to regard the principal goal of sample design as that of achieving a stated degree of precision for minimum cost, or alternatively, of maximizing precision for fixed cost.

Sample surveys in which major emphasis is placed on the estimation of population parameters such as percentages, means, or totals have been variously called "descriptive" or "enumerative" surveys and, as noted above, the work in finite-population sampling theory has been primarily concentrated on the design of such surveys. Increasingly, however, one finds reference in the sample survey literature to "analytical" surveys or to the use of "analytical statistics." Cochran (1963, p. 4), for example, says:

> In a descriptive survey the objective is simply to obtain certain information about large groups: for example, the numbers of men, women, and children who view a television program. In an analytical survey, comparisons are made between different subgroups of the population, in order to discover whether differences exist among them that may enable us to form or to verify hypotheses about the forces at work in the population. . . . The distinction between descriptive and analytical surveys is not, of course, clear-cut. Many surveys provide data that serve both purposes.

Although there are some differences in emphasis, Deming (1950, chap. 7), Hartley (1959), and Yates (1960, p. 297) make essentially the same distinction. Kish (1957, 1965) more or less automatically assumes that data derived from most complex sample surveys will be subjected to some type of detailed analysis, and applies the term "analytical statistical methods" to procedures that go much beyond the mere estimation of population percentages, means, and totals.

The Health Examination Survey (HES) of the National Center for Health Statistics (NCHS) is one example of a sample survey that, in some respects, might be classified as an enumerative survey, but whose principal value will undoubtedly be in providing data for analytical purposes. Some of the basic features of Cycle I of HES are presented in a publication of the National Center for Health Statistics (Series 11, No. 1). A brief description of the survey, quoted from the report, is as follows:

> The first cycle of the Health Examination Survey was the examination of a sample of adults. It was directed toward the collection of statistics on the medically defined prevalence of certain chronic diseases and of a particular set of dental findings and physical and physiological measurements. The probability sample consisted of 7,710 of all non-institutional, civilian adults in the age range 18-79 years in the United States. Altogether, 6,672 persons were examined during the period of the Survey which began in October 1959 and was completed in December 1962.

A rather detailed account of the survey design has been published by the Center (Series 1, No. 4).

The enumerative and analytical aspects of this survey, and the inevitable blending of one into the other, are well illustrated in two reports that have been published on the blood pressure of adults (NCHS, Series 11, Nos. 4 and 5). Not only does one find in these reports the distribution of blood pressure readings for the entire sample, but one also finds the comparison (with respect to blood pressure) of subgroups of the population defined by a variety of combinations of such demographic variables as age, sex, arm girth, race, area of the United States, and size of place of residence. It seems unnecessary to argue where the enumerative aspects end and the analytical aspects begin. For all practical purposes, and by any definition one chooses to adopt, the survey is analytical in character. The same will be true of almost any sample survey that one examines, at least as far as many users of the data are concerned. Certainly this is the view of the staff at NCHS.

The principal goal of this report will be to examine some of the problems that arise when data from a complex sample survey operation are subjected to detailed and critical analysis, and to discuss some of the procedures that have been suggested for dealing with these problems. Particular emphasis will be placed on a procedure for estimating variances which is especially suitable for sample designs similar to those used in the Health Interview Survey and the Health Examination Survey.

COMPLEX SAMPLE SURVEYS AND PROBLEMS OF CRITICAL ANALYSIS

Simple random sampling, usually without replacement, provides the base upon which the presently existing body of sample survey theory has been constructed. Major modifications of random sampling have been dictated by one or both of two considerations. These are as follows.

(1) One rarely attempts to survey a finite population without having some prior knowledge concerning either individual elements in the population, or subgroups of population elements, or the population as an entity. This prior information, depending upon its nature, can be used in the sample design or in the method of estimation to increase the precision of estimates over that which would be achieved by simple random sampling. Thus we have such techniques as stratification, stratification after the selection of the sample (poststratification), selection with probabilities proportional to the value of some auxiliary variable, ratio estimation, and regression estimation.

(2) Many finite populations chosen for survey study are characterized by one or both of the following two circumstances: the ultimate population elements are dispersed over a wide geographic area, and groups or "clusters" of elements can be readily identified in advance of taking the survey, whereas the identification of individual population elements would be much more costly. These circumstances have led to the use of multistage sampling procedures, where one first selects a sample of clusters and then selects a sample of elements from within each of the chosen clusters.

In addition to these two main streams of development, whose results are frequently combined in any one survey undertaking, there is a wide variety of related and special techniques from which choices can be made for sample design and for estimation. Thus one can use systematic sampling, rotation sampling (in which a population is sampled over time with some sample elements remaining constant from time to time), two-phase sampling (in which the results of a preliminary sample are used to improve design or estimation for a second sample), unbiased ratio estimators instead of ordinary ratio estimators, and so on. Finally, it is necessary to recognize that measurements may not be obtained from all elements that should have been included in a sample, and that such nonresponse may influence the estimation procedure and the interpretation of results.

As sample design and estimation move from simple random sampling and the straightforward estimation of population means, percentages, or totals to a stratified, multistage design with ratio or regression forms of estimation, it becomes increasingly difficult to operate in an "ideal" manner even for the purest of enumerative surveys. Ideally, one would like to be assured that the "best" possible estimate has been obtained for the given expenditure of funds, that the bias of the estimate is either negligible or measurable, and that the precision of the estimate has been appropriately evaluated on the basis of the sample selected. Numerous difficulties are encountered in achieving this goal. Among these are: (1) the expressions that must be evaluated from sample data become exceedingly complex, (2) in many instances, these expressions are only approximate in that their validity depends upon having "large" samples, and (3) most surveys provide estimates for many variables—that is, they are multivariate in character—and this, in conjunction with the first point, implies an extremely large volume of computations, even for modern electronic computing equipment. This last point is accentuated when one wishes to study the relationships among many variables in numerous subpopulations. The foregoing difficulties are well illustrated by the Health Interview Survey (No. A-2), the Health Examination Survey (Series 1, No. 4), and the Current Population Survey (Technical Paper No. 7). The sample designs and estimation techniques for these three surveys are somewhat similar, although the Current Population Survey employs a composite estimation technique (made possible by the rotation of sample elements) that is not employed in the other two surveys.

We have at various points in the preceding discussion used the term "complex" sample surveys, implying thereby that the sample design is in some sense or other complex. Little is to be gained by arguing the distinction between simple or complex under these circumstances, although several observations are perhaps in order. We are, of course, primarily concerned with the complexities of analysis that result from the use of a particular sample design and estimation procedure. These complexities arise from various combinations of such factors as the following. The assumption of a functional form for the distribution of a random variable over a finite population is rarely feasible and thus analytical power for devising statistical procedures is lost. The selection of elements without replacement, or in clusters, introduces dependence among observations. Estimators are usually nonlinear and we are forced to use approximate procedures for evaluating their characteristics. Some design techniques that are known to increase the precision of estimates almost invariably lead to the negation of assumptions required by such common statistical procedures as the analysis of variance (e.g., strata having unequal within-variances). Further comments on these points will be made later in this report.

New dimensions of complexity, both conceptual and technical, arise as one progresses from a truly enumerative survey situation to a purely analytical survey. Each of these will now be discussed briefly.

On the conceptual side, the major question concerns the manner in which one chooses to view a finite population—either as a fixed set of elements for which a statistical description is desired, or as a sample from an infinite superpopulation to which inferences are to be made. In simplest terms, this can be viewed as follows. An infinite superpopulation, characterized by random variable y with mean

$E(y) = \mu$, and with variance
$E(y-\mu)^2 = \sigma^2$,

is assumed as a basis for the sampling process. N independent observations on y lead to a finite population with mean

$(1/N)\sum_{i=1}^{N} y_i = \bar{Y}$, and with variance

$(1/N-1)\sum_{i=1}^{N}(y_i-\bar{Y})^2 = S^2$, while a simple random sample, drawn without replacement, from the finite population has the observed mean

$$(1/n)\sum_{i=1}^{n} y_i = \bar{y} \text{ and variance}$$
$$(1/n-1)\sum_{i=1}^{n}(y_i-\bar{y})^2 = s^2.$$

Ordinary sampling theory assumes that we wish to describe the realized finite population of N elements, and we have

$$E(\bar{y}|\text{fixed } N \text{ values of } y) = \bar{Y}$$
$$V(\bar{y}|\text{fixed } N \text{ values of } y) = \frac{N-n}{N}\frac{S^2}{n}$$
$$\hat{V}(\bar{y}|\text{fixed } N \text{ values of } y) = \frac{N-n}{N}\frac{s^2}{n}$$

where the symbol ^ indicates an estimator of a population parameter. If, however, we wish to draw inferences for the infinite superpopulation from our observed sample, and therefore take expectations over an infinite set of finite populations of N elements, then it is straightforward to demonstrate that

$$E(\bar{y}) = \mu$$
$$V(\bar{y}) = \sigma^2/n$$
$$\hat{V}(\bar{y}) = s^2/n$$

In effect, the only formal difference in the two views is that the finite population correction is omitted in the variance of $\bar{y}$ and in the estimate of the variance of $\bar{y}$. This point has been made by Deming (1950, p. 251) and Cochran (1963, p. 37). Cochran says, in reference to the comparison of two subpopulation means:

> One point should be noted. It is seldom of scientific interest to ask whether $\bar{Y}_j = \bar{Y}_k$ because these means would not be exactly equal in a finite population, except by a rare chance, even if the data in both domains were drawn at random from the same infinite population. Instead, we test the null hypothesis that the two domains were drawn from *infinite* populations having the same mean. Consequently we omit the *fpc* when computing $V(\bar{y}_j)$ and $V(\bar{y}_k)$. . . .

Actually, it can also be argued that one would rarely expect to find two infinite populations with identical means. Careful accounts of statistical inference sometimes emphasize this fact by distinguishing between "statistically significant difference" and "practically significant difference," and by pointing out that null hypotheses are probably never "exactly" true.

In practice, the survey sampler is ordinarily in a position to control *only* the inference from the sample to the finite population. He may know that the finite population is indeed a sample, *drawn in some completely unknown fashion* from an infinite superpopulation, but when he tries to specify this superpopulation, his definition will ordinarily be blurred and indistinct. Professional knowledge and judgment will therefore play a major role in such further inferences. Furthermore, comparisons with other studies, comparisons among subgroups in his finite population, and a consideration of related data must be brought to bear on the problem. There seems to be little that one can say in a definite way at present about this general problem, but very perceptive comments on this subject have been made by Deming and Stephan (1941) and by Cornfield and Tukey (1956, sec. 5). Even the answer to the specific question of whether or not to use finite population corrections in the comparison of domain means would appear to depend upon the circumstances.

The technical problems raised by the analytical use of data from complex surveys differ in degree but not really in kind from those faced in the consideration of enumerative survey data. These problems are primarily of two types:

1. As indicated earlier, most analytical uses of survey data involve the comparison of subgroups of the finite population from which the sample is selected. These subgroups have been frequently referred to as "domains of study." The basic difficulty raised by this fact is that various sample sizes, which in ordinary sam-

pling theory would be regarded as fixed from sample to sample, now become random variables. Furthermore, this occurs in such a manner that it is usually not possible to use a conditional argument—that is, it is not possible to consider the drawing of repeated samples in which the various sample sizes are viewed as being equal to the size actually observed—as can be done when estimating the mean of a domain on the basis of simple random sampling.

2. In making critical analyses of survey data, one is much more apt to use statistical techniques that go beyond the mere estimation of population means, percentages, and totals (e.g., multiple regression). Ordinary survey theory has attacked the problem of providing estimates of sampling error for certain estimates, e.g., ratio and regression estimates, but the body of available theory leaves much to be desired. An example of each of these problems will now be described briefly.

One of the most frequently used techniques from sampling theory is that of stratification. A population is divided into L mutually exclusive and exhaustive strata containing $N_1, N_2, \ldots, N_L$ elements; random samples of predetermined size $n_1, n_2, \ldots, n_L$ are drawn from the respective strata; a value of the variable y is obtained for *each* of the sample elements; and the population mean is estimated by

$$\hat{\bar{Y}} = \sum_{h=1}^{L} W_h \bar{y}_h ,$$

where $\bar{y}_h$ is the mean of the n_h elements drawn from the hth stratum, N is the total size of the population, and $W_h = N_h/N$. It is easily shown that an unbiased estimate of the variance of $\hat{\bar{Y}}$ is given by

$$\hat{V}(\hat{\bar{Y}}) = \sum_{h=1}^{L} W_h^2 (1-f_h)\, s_h^2/n_h$$

where s_h^2 is the variance for the variable y as estimated in the hth stratum, and $f_h = n_h/N_h$ is the sampling fraction in the hth stratum. For purposes of illustration, let us assume that the strata are geographic areas of the United States, that n_h, N_h, and N refer to all adults (18 years of age and over) and that the variable y is blood pressure.

Suppose now that one wishes to estimate the average blood pressure for males in the 40-45 year age range with arm girth between 38 and 40 centimeters. This special group of adults is a subpopulation, or domain of study, with reference to the total finite population, and elements of the domain will be found in each of the defined strata. The weights for the strata and the fixed sample sizes do not refer to this subpopulation and, over repeated drawings of the main sample, the number of domain elements drawn from a stratum will be a random variable. Furthermore, the total number of domain elements in a stratum is unknown. Under these circumstances, let

$n_{h,d}$ = the number of sample elements in the hth stratum falling in domain d.

$y_{hi,d}$ = the value of the variable for the ith sample element from domain d in the hth stratum.

$\hat{N}_d = \sum_{h=1}^{L} (1/f_h)\, n_{h,d}$ = the estimated total number of elements in the domain.

$\bar{y}_{h,d} = (1/n_{h,d}) \sum_{i=1}^{n_{h,d}} y_{hi,d}$ = sample mean, for hth stratum, of elements falling in domain d.

Then an estimate of the domain mean and its estimated variance are given by

$$\hat{\bar{Y}}_d = \frac{\sum_h (1/f_h) \sum_i y_{hi,d}}{\sum_h (1/f_h)\, n_{h,d}}$$

$$\hat{V}(\hat{\bar{Y}}_d) \doteq \frac{1}{\hat{N}_d^2} \sum_h \frac{N_h^2 (1-f_h)}{n_h(n_h-1)} \left[\sum_i (y_{hi,d} - \bar{y}_{h,d})^2 + n_{h,d}\left(1 - \frac{n_{h,d}}{n_h}\right)(\bar{y}_{h,d} - \hat{\bar{Y}}_d)^2 \right]$$

where $h = 1, 2, \ldots, L$ and $i = 1, 2, \ldots, n_{h,d}$.

These expressions have been presented and discussed by a number of authors—Durbin (1958, p. 117), Hartley (1959, p. 15), Yates (1960, p. 202), Kish (1961, p. 383), and Cochran (1963, p. 149. The factor $(1/\hat{N}_d^2)$ was evidently omitted in the printing of this formula).

Three points concerning these results are worthy of note in the context of the present discussion. These are:

1. The estimate is actually a ratio estimate—technically, a combined ratio estimate. It is therefore almost always biased for small sample sizes, and the variance formula is only approximately correct.

2. The complexity of the formulas, as regards derivation and computation, has been increased considerably over that of ordinary stratification.

3. The variance has a between-strata component as well as a within-strata component and, if $n_{h,d}$ is small as compared with n_h, this between-strata component can contribute substantially to the variance. Thus we see that changing emphasis from the total population to a subpopulation has introduced added complexities in theory and computations.

As regards the use of more advanced statistical techniques in the critical analysis of survey data, we shall simply refer to the difficulties that have been encountered in obtaining exact theory when ordinary regression techniques are applied to random samples drawn from a finite population. Cochran (1963, p. 193) summarizes this very well:

> The theory of linear regression plays a prominent part in statistical methodology. The standard results of this theory are not entirely suitable for sample surveys because they require the assumptions that the population regression of y on x is linear, that the residual variance of y about the regression line is constant, and that the population is infinite. If the first two assumptions are violently wrong, a linear regression estimate will probably not be used. However, in surveys in which the regression of y on x is thought to be approximately linear, it is helpful to be able to use $\bar{y}_{lr}$ without having to assume exact linearity or constant residual variance.
>
> Consequently we present an approach that does not demand that the regression in the population be linear. The results hold only in large samples. They are analogous to the large-sample theory for the ratio estimate.

Somewhat the same point is made by Hartley (1959, p. 24) in his paper on analyses for domains of study. He says:

> . . . nevertheless we shall *not* employ regression estimators. The reason for this is *not* that we consider regression theory inappropriate, but that this theory for finite populations requires considerable development before it can be applied in the present situation.

Some developments have arisen since Hartley's paper and reference to these will be given in the next section.

As a final complicating factor, we note that certain techniques used in some sample survey designs are such that their effects on the precision of estimates cannot be evaluated from a sample, even in the case of an enumerative survey. We refer specifically to the technique of controlled selection, described by Goodman and Kish (1950), and to instances in which only one first-stage sampling unit is selected from each of a set of strata (Cochran, 1963, p. 141).

In order to provide a convenient illustration of some of the foregoing points, the appendix presents a brief description of the "complex" sample design and estimation procedure employed in the Health Examination Survey, together with a selection of examples that arose in the more or less routine analysis of information collected on blood pressure. The data given are estimates of the percentage of individuals with hypertension in various subclasses of the population of adults, with the subclasses defined in terms of such demographic variables as race, sex, age, family income, education, occupation, and industry of employment. These subclasses cut across the strata used in the selection of primary sampling units, and the variances of the estimates are also affected by the various clustering and estimation features of the design.

Most of the cited cases refer to the estimation of variance for the percentage of hypertension in a single subclass, although several examples are given in which the percentages in two subclasses are compared. The variances were estimated by a replication technique that will be introduced in "Balanced Half-Sample Replication," a technique that to some extent overcomes the problems of analysis that have just been raised. The results obtained through the application of this technique will be used for illustration at several points throughout this report.

GENERAL APPROACHES FOR SOLVING PROBLEMS OF CRITICAL ANALYSIS

Two Extreme Approaches

It is possible to identify two extreme views that one may hold with respect to the problems raised in the preceding section. First of all, one might conceivably argue that analytical work with survey data should be done only "by design." That is, areas and methods of analysis should be set forth in advance of taking the survey and the sample should be selected so as to conform as closely as possible to the requirements of the stated methods. On the other extreme, one might decide to throw up his hands in dismay, ignore all the complicating factors of an already executed survey design, and treat the observations as though they had been obtained by random sampling, presumably from some extremely ill-defined superpopulation.

The first approach, that of "design for analysis," is certainly the most rational view that one can adopt. No careful examination of the literature was made to search out actual experiences on this point, but it would appear unlikely that one could find any examples of large-scale, complex, and multipurpose surveys in which this approach had been attempted. A possible exception might be the Census Enumerator Variation Study of the 1950 U.S. Census, as described by Hanson and Marks (1958), although this study was based primarily on the complete enumeration of designated areas rather than on a sample of individuals selected in accordance with a complex sample design. In other cases individuals have been randomly selected from a defined population of adults to provide observations for a complex "experimental design," as in the Durbin and Stuart (1951) experiment on response rates of experienced and inexperienced interviewers, but again this differs considerably from the type of problem raised in the preceding section. Another example in which an experimental design has been applied to survey data is provided by Keyfitz (1953) and discussed by Yates (1960, pp. 308-314). In this case, the sample elements were obtained by cluster sampling, but the author investigates the possible effects of the clustering and concludes that it can be ignored in the analysis of variance.

Some recent work by Sedransk (1964, 1964a, 1964b) bears directly on the problem of design for analysis and it assumes that the primary goal of an analytical survey is to compare the means of different domains of study. If $\hat{\bar{Y}}_i$ and $\hat{\bar{Y}}_j$ are the estimated means for the ith and jth domains, Sedransk places constraints on the variance of their difference, for all i and j, and searches for sample-size allocations that will minimize simple cost functions. A variety of different situations are considered. Random samples can be selected from each of the domains; random samples can be selected from the overall population, but the number of elements falling in each domain then becomes a random variable; two-stage cluster samples can be selected from each of the domains; and two-stage cluster samples can be selected from the total population, but again the number of elements falling in each domain is a random variable. In the second and fourth cases, the author considers double sampling procedures and obtains approximate solutions to guide one in choosing sample sizes for sampling from the total population so as to satisfy the constraints which are phrased in terms of all possible pair-wise domain comparisons. Even if one does not wish to impose constraints on domain comparisons and on minimization of cost, the cited papers contain of necessity many developments in theory that will be of assistance in attacking the problems raised in the preceding section. It should be observed that the complexity of the designs considered is still far from that of the Current Population Survey or the Health Examination Survey.

Major difficulties in designing for analysis are and will continue to be encountered when the primary goal of a survey is to describe a large and dispersed population with respect to many variables, as the analytical purposes are somewhat ill defined at the design stage. Thus the broad primary purposes of HES were to provide statistics on the medically defined prevalence in the total U.S. population of a variety of specific diseases, using standardized diagnostic criteria; and to secure distributions of the general population with respect to certain physical and physiological measurements. Nevertheless, analysis of relationships among variables is also an important product of the survey.

A similar set of circumstances arises with respect to data on unemployment collected by the Current Population Survey. Clearly the primary goal is to describe the incidence of employment and unemployment in the total U.S. population, and yet the data obtained must also be used for comparison and analysis. Faced with difficulties of analysis, as described in the preceding sections, one may wish to retreat to the opposite extreme from design for analysis and view the observations as coming from a simple random sample.

Actually, this type of retreat would appear to place the analyst in a difficult, if not untenable, position. Cornfield and Tukey (1956) speak of an inference from observations to conclusions as being composed of two parts, where the first part is a statistical bridge from observations to an island (the island being the studied population) and the second part is a subject-matter span from the island to the far shore (this being, in some vague sense, a population of populations obtained by changes in time, space, or other dimensions). The first bridge is the one that can be controlled by the use of proper procedures of sampling and of statistical inference. One may be willing to introduce some uncertainty into the position of the island, for example by ignoring finite population corrections, in the hope of placing it nearer the far shore than would otherwise be the case. However, there seem to be no grounds for suggesting statistical procedures that may, unbeknownst to their user, succeed only in moving the island a short distance from the near shore.

The Health Examination Survey was carried out on a sample chosen to be broadly "representative" of the total U.S. population. Among other characteristics of design, the sample was of necessity a highly clustered one. As is well known, a highly clustered sample leads to estimates that have much larger standard errors than would be predicted on the basis of simple random sampling theory if the elements within clusters tend to be homogeneous with respect to the variables of interest. Since geographic clustering leads to homogeneity on such characteristics as racial background, socioeconomic status, food habits, availability and use of medical care, and the like, it can therefore be expected that there will also be homogeneity with respect to many of the variables of interest in the Health Examination Survey. A portion of this loss of precision is undoubtedly recovered by stratification and poststratification, but there is no guarantee that the two effects will balance one another. Hence the ignoring of sample design features might well lead to gross errors in determining the magnitude of standard errors of survey estimates. In effect, the situation might be viewed as one in which inferences are being made to some ill-defined population of adults, rather than to the population from which the sample was so carefully chosen. These points have been emphasized by Kish (1957, 1959) as they relate to social surveys in general.

The effect of these sample design and estimation features on the variances of estimates for the Health Examination Survey are illustrated in the data presented in the appendix. For each of 30 designated subclasses, the variance of the estimated percentage of adults in a subclass with hypertension was estimated by two methods: (1) the replication technique to be introduced in "Balanced Half-Sample Replication" was employed, thus accounting for most of the survey features, and (2) the observations falling in a subclass were treated as if they had arisen from simple random sampling. In the second case, variances were computed as pq/n, where p was the observed fraction of hypertensive individuals among the n sample individuals falling in a particular subclass. The ratios of the first of these variances to the second, for the 30 comparisons, ranged from .45 to 2.87, with an average value of 1.31; the ratios of the standard errors ranged from .67 to 1.69, with an average value of 1.12. These ratios probably overestimate slightly the true ratios, since the replication technique uses the method of collapsed strata and in this instance does not account for the effects of controlled selection. Furthermore, they are subject to sampling variability.

Also included in the appendix are three examples which refer to the estimated difference between the percentages of hypertensive individuals in each of two subclasses. In this instance, the average ratio of variances is 1.51 while the average ratio of standard errors is 1.23. These comparisons are not as "clean" as the ones for single subclasses since the random sampling variances were computed on the assumption of independence of the two estimates, and this is not necessarily the case. This set of data, limited though it may be, tends to confirm the general experience that estimates made from stratified cluster samples will tend to have larger sampling variances than would be the case for simple random samples of the same size, although the differences are not so pronounced as in situations in which the intraclass correlation is stronger than it is for this statistic.

Obtaining "Exact" Solutions

If one wishes to consider that the principal goal of analytical surveys is either the estimation or the direct comparison of the means of various domains of study, then there already exists in the literature a number of results that can assist in achieving this goal. This "exact" theory can be generally characterized as follows.

1. Ratio estimates of population means are employed, primarily because sample size is a random variable as a result of sampling clusters with unequal and unknown sizes. This of course introduces the possibility of bias for these estimates, although empirical research—e.g., Kish, Namboodiri, and Pillai (1962)—indicates that the amount of bias is apt to be negligible.

2. Expressions for the variance of a single estimate and the covariances of two or more estimates are obtained from the Taylor series approximation, and variance estimates are constructed by direct substitution into these expressions. Hence variance estimates are subject to possible bias.

3. In multistage sampling, it is either assumed that the first-stage units are drawn with replacement or that the first-stage sampling fractions are very small. This means that variance estimates can be obtained without explicitly treating within-first-stage unit sampling variability.

4. The most powerful tool for deriving results for domain-of-study estimates has been that of the "pseudovariable" and the "count variable." That is

$y'_{hij} = y_{hij}$, if the jth element in the ith first-stage unit of the hth stratum belongs to the domain of interest

$= 0$, otherwise

$u_{hij} = 1$, if the jth element in the ith first-stage unit of the hth stratum belongs to the domain of interest

$= 0$, otherwise

Using this approach, which is related to Cornfield's (1944) earlier work, it is possible to specialize ordinary results to domain-of-study results.

A very brief summary of some of the literature on these aspects of analysis is as follows, where no attempt has been made to assign priorities to the various authors. Results that are directly phrased in terms of domains of study are given by Cochran (1963), Durbin (1958), Hartley (1959), Kish (1961, 1965), and Yates (1960). Related work on the estimation of the variance of a variety of functions of ratio estimators is presented by Jones (1956), Keyfitz (1957), Kish (1962), and Kish and Hess (1959). Aoyama (1955), Garza (1961), and Okamoto (1963) discuss chi-square contingency-table analyses in the presence of stratification. McCarthy (1965) considers the problem of determining distribution-free confidence intervals for a population median on the basis of a stratified sample. Finally, we observe that, in the case of "small" samples, not even the ordinary normality assumptions are able to do away with difficulties, even without domain-of-study complications. Thus unequal strata variances lead to difficulties in obtaining tests and confidence intervals for a population mean, although some approximate solutions are available—e.g., Aspin (1949), Meier (1953), Satterthwaite (1946), and Welch (1947). If one has essentially unrecognized stratification—that is, normal variables with common variance but differing means—then it is necessary to work with noncentral t, χ^2, and F distributions as described by Weibull (1953).

Replication Methods of Estimating Variances

As a result of the indicated theoretical and practical difficulties associated with the estimation of variances from complex sample surveys, interest has long been evidenced in developing shortcut methods for obtaining these estimates. For example, we noted earlier that Keyfitz (1957) and Kish and Hess (1959) have emphasized the computational simplicity that can result when primary sampling units are drawn with replacement from each of a number of strata, and when one can work with variate values associated with the primary sampling units. There are, however, other approaches that have been suggested and applied to accomplish these same ends. We refer in particular to methods that have variously been referred to as interpenetrating samples, duplicated samples, replicated samples, or random groups. In the succeeding discussion, the term "replicated sampling" will be used to cover all of these possibilities. References will be made to the pertinent literature, but no attempt will be made to assign priorities or to be exhaustive. Deming has been a consistent and firm advocate of replicated sampling. He first wrote of it as the Tukey plan (1950); his recent book (1960) presents descriptions of the applications of replicated sampling to many different situations, and contains a wide variety of ingenious devices that he has developed for solving particular problems.

In simplest form, replicated sampling is as follows. Suppose one obtains a simple random sample of n observations—drawn with replacement from a finite population or drawn independently from an infinite population—and that the associated values of the variable of interest are $y_1, y_2, \ldots, y_n$. Then, if $\bar{y}$ denotes the sample mean and $\bar{Y}$ the population mean, $E(\bar{y}) = \bar{Y}$ and

$$\hat{V}_n(\bar{y}) = \sum_{i=1}^{n} (y_i - \bar{y})^2 / n(n-1)$$

$\hat{V}_n(\bar{y})$ provides an unbiased estimate of $V(\bar{y})$. Suppose now that n observations are randomly divided into t mutually exclusive and exhaustive groups, each containing (n/t) elements, and that the means of these groups are denoted by $\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_t$. It is clear that

$$\bar{y} = \sum_{j=1}^{t} \bar{y}_j / t$$

and that the variance of $\bar{y}$ can be estimated by

$$\hat{V}_t(\bar{y}) = \sum_{j=1}^{t} (\bar{y}_j - \bar{y})^2 / t(t-1)$$

In this simple case, the advantages gained by using $\hat{V}_t(\bar{y})$ rather than $\hat{V}_n(\bar{y})$ lie in the fact that one has to compute the sum of t squared deviations instead of the sum of n squared deviations. If t is considerably smaller than n and if such computations must be carried out for many variables, the savings in computational time may be substantial. Also, the kurtosis of the distribution of the $\bar{y}_i$ is less than that of y_i, possibly offsetting some of the effect of having fewer degrees of freedom to estimate $V(\bar{y})$.

There is, of course, a loss of information associated with the subsample approach for estimating variances since $\hat{V}_t(\bar{y})$ is subject to greater sampling variability than is $\hat{V}_n(\bar{y})$. A variety of ways have been suggested for measuring this loss of information. Hansen, Hurwitz, and Madow (1953, vol. I, pp. 438-449), who designate this the random group method of estimating variances, make the comparison in terms of the relative-variance of the variance estimate. For example, they show by way of illustration that the relative-variance of a variance estimate based on 1,200 observations drawn from a normal distribution is 4.1 percent while the relative-variance based on a sample of 60 random groups of 20 observations each is 18.3 percent. Actually, this approach places emphasis on the variance estimate itself rather than on the fact that one usually wants to use the variance estimate in setting confidence limits for a population mean or in testing hypotheses about a population mean. Under these circumstances, Fisher (1942, sec. 74) suggests a measure for the amount of information that a sample mean provides respecting a population mean. Since his approach has been questioned by numerous authors—e.g., Bartlett (1936)—we shall simply adopt the expedient of taking the ratio of, say, the 97.5 percentiles of Student's t distribution for 1,200 and 60 degrees of freedom, which is .981, and interpreting this as a measure of the relative width of the desired confidence intervals. This is evidently the approach used by Lahiri (1954, p. 307).

The foregoing is, of course, the familiar argument that a sample of roughly 30 or more is a "large" sample when dealing with normal populations, since $t_{.975}$ for 30 degrees of freedom is 2.042 and for a normal distribution is 1.960. Ninety-five percent confidence intervals will, on the average, be only about four percent wider when s^2 is estimated with 30 degrees of freedom than when σ^2 is known. A slightly different measure has been proposed by Walsh (1949), formulated in terms of the power of a t-test.

If one wishes to use replicated sampling in conjunction with drawing *without replacement* from a finite population, then two different possibilities arise. One can first draw without replacement a sample of (n/t) elements, then replace these elements in the population and draw a second sample of (n/t) elements, and continue this process until t samples have been selected. Denoting the sample means by $\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_t$, we have

$$\bar{y} = \sum_{j=1}^{t} \bar{y}_j / t$$

$$\hat{V}_t(\bar{y}) = \sum_{j=1}^{t} (\bar{y}_j - \bar{y})^2 / t(t-1)$$

and

$$E[\hat{V}_t(\bar{y})] = \frac{N-(n/t)}{N} \frac{S^2}{t(n/t)} = \frac{N-(n/t)}{N} \frac{S^2}{n}$$

This type of replication makes the successive samples independent of one another, but it does permit the possible duplication of elements in successive samples and hence lowers the precision of $\bar{y}$ as compared with the original drawing of a sample of n elements without replacement. However, there may be a saving in the cost of measuring the duplicated items. Hence a slightly larger sample could be drawn for the same total cost, recapturing some of the loss of precision.

Finally, one can draw a sample of (n/t) elements without replacement, a second sample without replacing the first sample, and so on. This is, of course, equivalent to drawing a sample of n elements without replacement and then randomly dividing the sample into t groups. It follows that

$$\bar{y} = \frac{\sum_{j=1}^{t} \bar{y}_j}{t}$$

$$\hat{V}_t(\bar{y}) = \frac{N-n}{N} \frac{\sum_{j=1}^{t} (\bar{y}_j - \bar{y})^2}{t(t-1)}$$

and

$$E[\hat{V}_t(\bar{y})] = \frac{N-n}{N} \frac{S^2}{n}$$

This latter variance is smaller than the preceding one because of the difference in finite population corrections. Note, however, that the independence achieved by the first method of drawing makes it easy to apply nonparametric methods (e.g., the use of order statistics) for estimation and hypothesis testing. These points have been discussed, in a more general framework, by Koop (1960) and by Lahiri (1954).

Although replicated sampling for simple random sampling, as described in the preceding paragraphs, does provide the possibility of achieving some gains in terms of computational effort, the principal advantages of replication arise from other facets of the variance estimation problem. Some of these facets may be identified as follows.

1. There are instances of sample designs in which no estimate of sampling precision can be obtained from a single sample unless certain assumptions are made concerning the population. Systematic sampling is a case in point. (See

Cochran, 1963, pp. 225,226.) If the total sample is obtained as the combination of a number of replicated systematic samples, then one can obtain a valid estimate of sampling precision. This approach was suggested by Madow and Madow (1944, pp. 8,9) and has been discussed at greater length and with a number of variations by Jones (1956). In some instances, estimates made from replicated systematic samples may be less efficient than from a single systematic sample, and then one must choose between loss of efficiency and ease of variance estimation, as discussed by Gautschi (1957).

2. As is well known, the ordinary Taylor series approximation for obtaining the variance and the estimated variance of a ratio estimate, even for simple random sampling, provides a possibly biased estimate of sampling precision. As an alternative, one can consider drawing a number of independent random samples, computing a ratio estimate for each sample, and then averaging these ratio estimates for the final estimate. A valid estimate of sampling precision can then be obtained from the replicated values of the estimate. It is true, however, that the bias of the estimator itself is undoubtedly larger for the average of the separate estimates than it is for a ratio estimate computed for the complete sample since this bias is ordinarily a decreasing function of sample size. Thus gains may be achieved in one respect, while losses may be increased in the other. As far as the author knows, no completed research is available to guide one in making a choice between these two specific alternatives. This problem is, however, related to some suggestions and work by Mickey, Quenouille, Tukey, and others, and their results will be discussed in some detail in the following section.

3. After an estimate and an estimated variance have been obtained, confidence intervals are ordinarily set by appealing to large sample normality and to the approximate validity of Student's t distribution. Replication can sometimes assist in providing "better" solutions. For example, consider a stratified population in which the variable of interest has a normal distribution within each stratum, but where the variance is different for the separate strata. Difficulty is then encountered in applying the chi-square distribution to the ordinary estimate of variance, as discussed by Satterthwaite (1946), Welch (1947), Aspin (1949), and Meier (1953). However, the mean of a replicate will be normally distributed, being a linear combination of normally distributed variables, and the chi-square distribution can be applied directly to a variance estimated from the means of a number of independent replicates. This aspect of the problem has been discussed at some length by Lahiri (1954, p. 309).

4. If one is using a highly complex sample design and estimation procedure, and if independent replicates can be obtained, then replicated sampling permits one to bypass the extremely complicated variance estimation formulas and the attendant heavy programming burdens. Variance estimates based upon the replicated estimates will mirror the effects of all aspects of sampling and estimation that are permitted to vary randomly from replicate to replicate. This, of course, includes the troublesome domain-of-study type of problem.

One major disadvantage of replicated sampling has been mentioned in the preceding paragraphs, namely that the variance estimate refers to the average of replicate estimates rather than to an estimate prepared for the entire sample. If the estimates are linear in the individual observations, the two will be the same. They will not be the same, however, for the frequently employed ratio estimator and the other nonlinear estimators, and the average of the replicate estimators may possibly be subject to greater bias than is the case for the overall sample estimate.

Another major disadvantage arises from the difficulty of obtaining a sufficient number of replicates to provide adequate stability for the estimated variance. Thus the commonly used design of two primary sampling units per stratum (frequently obtained by collapsing strata from each of which only a single unit has been drawn) gives only two independent replicates, and the resulting confidence intervals for an estimate are much wider than they should be or need to be. Some suggestions have been made for attacking this problem, and they will be discussed in the following section.

Another, but subsidiary, problem arises with replication if one wishes to estimate components of variance—that is, to determine what fraction of the total variance of an estimate arises from the sampling of primary sampling units, what fraction arises from sampling within primary sampling units, and the like. This problem does not appear to have been discussed at any great length in the sampling literature and will not be considered here since it bears more directly on design than on analysis. Some of Sedransk's work (1964, 1964a, and 1964b) does relate to the problem, and McCarthy (1961) has discussed the matter in connection with sampling for the construction of price indexes.

PSEUDOREPLICATION

Half-Sample Replication Estimates of Variance From Stratified Samples

If a set of primary sampling units is stratified to a point where the sample design calls for the selection of two primary sampling units per stratum, there are only two independent replicates available for the estimation of sampling precision. Confidence intervals for the corresponding population parameter will then be much wider than they need to be. To overcome this difficulty, at least partially, the U.S. Bureau of the Census originated a pseudoreplication scheme called half-sample replication. The scheme has been adapted and modified by the NCHS staff and has been used in the HES reliability measurements. A brief description of this approach is given in a report of the U.S. Bureau of the Census (Technical Paper No. 7, p. 57), and a reference to the Census Bureau method of half-sample replication was made by Kish (1957, p. 164). We shall first present a technical description of half-sample replication as used by NCHS in the Health Examination Survey. The theoretical development duplicates, in part, work by Gurney (1962). We shall then suggest several ways in which the method can be modified to increase the precision of variance estimates.

Consider a stratified sampling procedure where two independent selections are made in each stratum. Let the population and sample characteristics be denoted as follows:

Stratum	*Weight*	*Population mean*	*Population variance*	*Sample observations*	*Sample mean*
1	W_1	$\bar{Y}_1$	S_1^2	y_{11}, y_{12}	$\bar{y}_1$
2	W_2	$\bar{Y}_2$	S_2^2	y_{21}, y_{22}	$\bar{y}_2$
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
h	W_h	$\bar{Y}_h$	S_h^2	y_{h1}, y_{h2}	$\bar{y}_h$
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
L	W_L	$\bar{Y}_L$	S_L^2	y_{L1}, y_{L2}	$\bar{y}_L$
	1				

An unbiased estimate of the population mean $\bar{Y}$ is $\bar{y}_{st} = \sum_{h=1}^{L} W_h \bar{y}_h$, and the ordinary sample estimate of $V(\bar{y}_{st})$ is

$$v(\bar{y}_{st}) = (1/2)\sum_{h=1}^{L} W_h^2 s_h^2 = (1/4)\sum_{h=1}^{L} W_h^2 d_h^2$$

where $d_h = (y_{h1} - y_{h2})$.

Under these circumstances, a half-sample replicate is obtained by choosing one of y_{11} and y_{12}, one of y_{21} and $y_{22}, \ldots,$ and one of y_{L1} and y_{L2}. The half-sample estimate of the population mean is

$$\bar{y}_{hs} = \sum_h W_h y_{hi}$$

where i is *either* one or two for each h. There are 2^L possible half samples, and it is easy to see that the average of *all* half-sample estimates is equal to $\bar{y}_{st}$. That is, for a randomly selected half sample

$$E(\bar{y}_{hs} | y_{11}, y_{12}, \ldots, y_{L1}, y_{L2}) = \bar{y}_{st}$$

If one considers the deviation of the mean determined by a particular half sample, for example $\bar{y}_{hs,1} = \Sigma W_h y_{h1}$, from the overall sample mean, the result is obtained that

$$(\bar{y}_{hs,1} - \bar{y}_{st}) = \sum_h W_h y_{h1} - (1/2) \sum_h W_h (y_{h1} + y_{h2})$$

$$= (1/2) \Sigma W_h (y_{h1} - y_{h2}) = (1/2) \Sigma W_h d_h$$

In general, these deviations are of the form

$$(\bar{y}_{hs} - \bar{y}_{st}) = (1/2)(\pm W_1 d_1 \pm W_2 d_2 \dots \pm W_L d_L)$$

where the deviation for a particular half sample is determined by making an appropriate choice of a plus or minus sign for each stratum. In the example given above, each sign was taken as plus. The squared deviation from the overall sample mean is therefore of the general form

$$(\bar{y}_{hs} - \bar{y}_{st})^2 = (1/4) \Sigma W_h^2 d_h^2 + (1/2) \sum_{h<k} \pm W_h W_k d_h d_k \quad (4.1)$$

where the plus or minus signs in the cross-product summation are determined by the particular half sample that is used.

If the squared deviations of a half-sample estimate from the overall sample mean are summed over all possible half samples, then it is easy to demonstrate that the cross-product terms appearing in the separate squared deviations cancel one another. Thus, for a randomly selected half sample

$$E[(\bar{y}_{hs} - \bar{y}_{st})^2 | d_1, d_2, \dots, d_L] = (1/4) \Sigma W_h^2 d_h^2 = v(\bar{y}_{st})$$

Since $v(\bar{y}_{st})$ is known to be an unbiased estimate of $V(\bar{y}_{st})$ if one takes expected values over repeated selections of the entire sample, we have the result that

$$E(\bar{y}_{hs} - \bar{y}_{st})^2 = (1/2) \Sigma W_h^2 S_h^2 = V(\bar{y}_{st}) \quad (4.2)$$

This also follows directly from (4.1) if we note that $E(d_h d_k) = 0$ because of the independence of selections within strata. If the sampling fractions are the same in all strata, then a finite population correction can easily be inserted at the end of the variance estimation process. The effect of differing sampling fractions could be taken into account by working with W_h' 's, where W_h' is equal to W_h multiplied by the square root of the appropriate finite population correction.

The foregoing properties of a half-sample estimate of the population mean have been exploited for variance estimation in the following manner.

Consider the population of 2^L possible half-sample estimates, for fixed values of $y_{11}, y_{12}, \dots, y_{L1}, y_{L2}$. This population has mean equal to $\bar{y}_{st}$. Furthermore, the mean value of $(\bar{y}_{hs} - \bar{y}_{st})^2$ is equal to $v(\bar{y}_{st})$. Draw with replacement, a sample of k half samples. Denote their means by $\bar{y}_{hs,1}, \bar{y}_{hs,2}, \dots, \bar{y}_{hs,k}$. Then take

$$v_{hs,k}(\bar{y}_{st}) = \sum_{i=1}^{k} (\bar{y}_{hs,i} - \bar{y}_{st})^2 / k \quad (4.3)$$

as an estimate of $V(\bar{y}_{st})$. Since the expected value of each squared term is equal to $V(\bar{y}_{st})$, the expected value of their average is also equal to $V(\bar{y}_{st})$. The different squares are not independent, since the $\bar{y}_{hs,i}$'s have elements in common, but this does not influence the expected value.

Quite clearly, $v_{hs,k}(\bar{y}_{st})$ is only an approximation to $v(\bar{y}_{st})$ which, in turn, is only an approximation to $V(\bar{y}_{st})$. Let us therefore attempt to estimate the degree of approximation by computing the variance of $v_{hs,k}(\bar{y}_{st})$. The relationship given in Hansen, Hurwitz, and Madow (1953, vol. II, p. 65)

$$V(u) = E_v[V(u|v)] + V_v[E(u|v)] \quad (4.4)$$

will be applied to the variable $u = (\bar{y}_{hs} - \bar{y}_{st})^2$, and the conditioning variable v will be replaced by the set of d_h 's. The first term then becomes

$$E_{d_1 \dots d_L}[V\{(\bar{y}_{hs} - \bar{y}_{st})^2 | d_1 \dots d_L\}]$$

By definition, where it is understood that all d_h 's are fixed,

$$V[(\bar{y}_{hs} - \bar{y}_{st})^2] = E\left[(\bar{y}_{hs} - \bar{y}_{st})^2 - (1/4) \Sigma W_h^2 d_h^2\right]^2$$

$$= \frac{1}{2^L} \sum_{i=1}^{2^L} \left[(1/2) \sum_{h<j} a_{hj,i} (W_h d_h)(W_j d_j)\right]^2$$

where the summation on i is over all possible half samples, and $a_{hj,i}$ is either plus or minus one, depending on the particular half sample and the particular combination of h and j. This expression is equal to

$$\frac{1}{2^{L+2}}\sum_{i=1}^{2^L}\left[\sum_{h<j}(W_h d_h)^2(W_j d_j)^2 + \text{(sum of cross-product terms, each term containing at least one } d_j \text{ to the first power)}\right]$$

It can be demonstrated that the cross-product terms, when summed over all 2^L samples will vanish. Hence

$$V[(\bar{y}_{hs}-\bar{y}_{st})^2 | d_1 \ldots d_L]$$

$$= (1/4) \sum_{h<j} (W_h d_h)^2 (W_j d_j)^2$$

Since $E(d_h^2/2) = S_h^2$ and d_h and d_j are independent, the expectation of this variance over sets of d_h is equal to

$$\sum_{h<j}(W_h^2 S_h^2)(W_j^2 S_j^2) \qquad (4.5)$$

We shall now evaluate the second term in (4.4). Since

$$E[(\bar{y}_{hs}-\bar{y}_{st})^2 | d_1 \ldots d_L] = (1/4)\,\Sigma\, W_h^2 d_h^2$$

We have

$$V_v[E(u|v)] = E[(1/4)\,\Sigma\, W_h^2 d_h^2 - (1/2)\,\Sigma\, W_h^2 S_h^2]^2$$

$$= (1/4) E\,[\Sigma\, W_h^2(\frac{d_h^2}{2} - S_h^2)]^2 \qquad (4.6)$$

$$= (1/4)\,\Sigma\, W_h^4\, E\,(\frac{d_h^2}{2} - S_h^2)^2$$

where the cross-product terms vanish because the selections within strata are independent from one stratum to another and because $E(d_h^2/2) = S_h^2$

Putting expressions (4.5) and (4.6) together, using relative variance (rel-variance) instead of absolute variance, and taking account of the fact that we are averaging k half-sample estimates of variance (which affects only the first term) we have

$$\text{Rel-var}\,[v_{hs,k}(\bar{y}_{st})] = \qquad (4.7)$$

$$[(4/k)\sum_{h<j}(W_h^2 S_h^2)(W_j^2 S_j^2) + \Sigma\, W_h^4\, E(\frac{d_h^2}{2} - S_h^2)]/(\Sigma\, W_h^2 S_h^2)^2$$

The exact value of this expression depends upon the values of $W_h^2 S_h^2$ and the distribution of the y_{ij} 's within strata. However, reasonable approximations can be obtained. Thus the first term in (4.7) is a maximum, for fixed value of $\Sigma\, W_h^2 S_h^2$, if $W_1^2 S_1^2 = W_2^2 S_2^2 \ldots = W_L^2 S_L^2 = W^2 S^2$. Under these circumstances, the value of this term is

$$\frac{4}{k}\,\frac{W^4 S^4}{W^4 S^4}\,\frac{\frac{L(L-1)}{2}}{L^2} = \frac{2\,(L-1)}{kL}$$

For the second term, assume that the strata weights are equal, that the strata variances are equal, and that $\beta (= \mu_4/\sigma^4)$ is the same for each stratum. Using the expression for the rel-variance of the estimated variance (Hansen, Hurwitz, and Madow, 1953, vol. II, p. 99) the second term in (4.7) is then approximated by

$$\frac{\beta+1}{2}\cdot\frac{1}{L}$$

Then the rel-variance of $v_{hs,k}(\bar{y}_{st})$ is approximated by

$$\frac{2(L-1)}{kL} + \frac{\beta+1}{2L} \qquad (4.8)$$

This expression, as earlier developed by Gurney (1962), was used in obtaining values given in a U.S. Census report (Technical Paper No. 7, table 29), where L was taken to be 85 and a table was prepared for several different values of β.

The Health Examination Survey is based on 42 strata, collapsed into 21 strata for the present model. Hence we have the results given in table 1 for $\beta = 3$ and $\beta = 6$.

Table 1. Rel-variance of a variance estimated from k independent half-sample replications

(Number of strata, L, is 21)

k	$\beta = 3$	$\beta = 6$
1	2.0000	2.0715
2	1.0476	1.1191
5	.4762	.5477
10	.2857	.3572
20	.1904	.2619
40	.1428	.2143
100	.1142	.1857
∞	.0952	.1667

Table 2. Equivalent degrees of freedom obtained from k independent half-sample replications

(Analysis of variance assumptions; 21 strata)

k	Rel-variance of estimated variance	Equivalent degrees of freedom	$t_{.975}$
1-----	2.0000	1.0	12.706
2-----	1.0476	1.9	4.740
5-----	.4762	4.2	2.727
10----	.2857	7.0	2.365
20----	.1904	10.5	2.213
40----	.1428	14.0	2.145
100---	.1142	17.5	2.105
∞ ---	.0952	21.0	2.080

Table 3. Relative width of confidence intervals obtained from k independent half-sample replications

(Analysis of variance assumptions; 21 strata)

k	Width of confidence interval relative to the full 21 degrees of freedom
1-----------------	6.109
2-----------------	2.279
5-----------------	1.311
10----------------	1.137
20----------------	1.064
40----------------	1.031
100---------------	1.012
∞ ----------------	1.000

If one is committed to using independent half-sample replications, to which the above table refers, an appropriate number of replicates must be chosen. In some sense or other, one would like to use a sufficient number of replicates so that not "too much" of the available information is lost. On the other hand, this desire must be balanced against the cost of processing "too many" replicates. If a simple analysis of variance model is assumed, it is possible to obtain some idea of the amount of information that is lost for various numbers of replicates. For example, assume that the strata are of equal weight, that the variable y has a normal distribution in each stratum, that the strata means are possibly different, and that the variances within strata have a common value S^2. Under these circumstances, the ordinary within-stratum estimate of S^2 has 21 degrees of freedom, and confidence intervals for a population mean would be based on Student's t distribution with 21 degrees of freedom.

For a normal distribution, the relative variance of an estimated variance is $2/(n-1)$—Cochran (1963, p. 43)—where n is the number of independent observations. If the relative variances given in table 1 for $\beta = 3$ are set equal to $2/(n-1)$ and we solve for $(n-1)$, the result can be viewed as "equivalent degrees of freedom" for use with Student's t distribution. The values obtained from this process are given in table 2.

As previously noted, there are a variety of ways that one can make comparisons among these various cases, but the simplest would appear to be on the basis of width of the 95 percent confidence interval, for fixed value of the estimated variance. This approach has evidently been used by Lahiri (1954, p. 307). Table 3 points this out. Remembering that the determination of the relative variance of the variance is probably an overestimate, and that the effect of replication is less pronounced for values of β greater than 3, it would appear that between 20 and 40 independent replications would work out quite satisfactorily in practice. This statement must of course be qualified in terms of the cost of additional replications.

Balanced Half-Sample Replication

Although the point has not been explicitly emphasized, the preceding development shows that the variability among half-sample estimates of variance arises from between-strata contributions to these estimates, as is evident from equations (4.1) and (4.5). These contributions come from the cross-product terms involving $d_h\ d_k$. These cross-product terms cancel one another over the entire set of 2^L half samples, or when one uses an infinite number of half-sample replications. The question then arises whether one can choose a relatively small subset of half samples for which these terms will also disappear. If this can be done, then the corresponding half-sample estimates of variance will contain all of the information (21 degrees of freedom in the simple situation being considered here) available in the total sample. Some attempt at "balancing" is indicated in the U.S. Bureau of the Census publication (Technical Paper No. 7, p. 57) and this is described in the following quotation:

> For the non-self-representing strata, the strata were grouped into 85 homogeneous pairs. For each of these 85 pairs, a set of 20 random numbers between 1 and 40 were selected without replacement. The first PSU in the pair was included in the 20 replications corresponding to these 20 random numbers, and the second PSU in the remaining 20 replications. In this way, a set of 85 PSU's, one from each pair, was assigned to each replication. It would have been possible to make an independent selection from each pair for each replication; however, it is believed that the reliability of the estimates was increased somewhat by insuring that each PSU would appear in 20 of the 40 replications.

It is unlikely, however, that this type of balancing will appreciably affect the reliability of variance estimates, since as noted above, this reliability is determined by the joint occurrence of elements from pairs of strata.

A simple example will show that it is possible to select a subset of half-sample replications that will have the desired property. Consider a three-strata situation with observations (y_{11}, y_{12}), (y_{21}, y_{22}), and (y_{31}, y_{32}). There are $2^3 = 8$ possible half-sample replicates. Now consider the following subset of four replicates:

Replicate	*Stratum* 1	2	3	$(\bar{y}_{hs,i} - \bar{y}_{st})$
1	y_{11}	y_{21}	y_{31}	$(1/2)(+W_1 d_1 + W_2 d_2 + W_3 d_3)$
2	y_{11}	y_{22}	y_{32}	$(1/2)(+W_1 d_1 - W_2 d_2 - W_3 d_3)$
3	y_{12}	y_{22}	y_{31}	$(1/2)(-W_1 d_1 - W_2 d_2 + W_3 d_3)$
4	y_{12}	y_{21}	y_{32}	$(1/2)(-W_1 d_1 + W_2 d_2 - W_3 d_3)$

The signs of the separate terms in the deviations are determined by the definition of $d_h = (y_{h1} - y_{h2})$. It is, of course, immaterial how the two observations within a stratum are numbered originally. Once the numbering is set, however, as in the first replicate, it is maintained in determining the remaining replicates. If these deviations are squared, the first part of each expression is $W_1^2 d_1^2/4 + W_2^2 d_2^2/4 + W_3^2 d_3^2/4$, which is the desired estimate of variance. The second part of each expression contains the cross-product terms, and it can easily be checked that all these cross-product terms cancel when the squared deviations are added over the four replicates. This follows from the fact that the columns of the matrix of signs in the deviations are orthogonal to one another. Thus this set of balanced half samples can be identified as

```
+ + +
+ - -
- - +
- + -
```

where a plus sign indicates y_{h1}, while a minus sign denotes y_{h2}. Notice that this particular set of replicates does have the property that each of the two elements in a stratum appears in half the samples.

If one wishes to obtain a set of half samples that will have this feature of "cross-product balance," for any fixed number of strata, then it becomes necessary to have a method of constructing matrices of + and - signs whose columns are orthogonal to one another. A method is described by Plackett and Burman (1943-46, p. 323) for obtaining $k \times k$ orthogonal matrices, where k is a multiple of 4. Suppose, for example, that we have 5, 6, 7, or 8 strata. The Plackett-Burman method produces the following 8 x 8 matrix, which is the smallest that can be used for these cases because of the multiple-of-4 restriction. The rows identify a half sample, while the columns refer to strata.

```
+ - - + - + + -
+ + - - + - + -
+ + + - - + - -
- + + + - - + -
+ - + + + - - -
- + - + + + - -
- - + - + + + -
- - - - - - - -
```

Any set of 5 columns for the 5 strata case, 6 columns for the 6 strata case, 7 columns for the 7 strata case, or the entire 8 columns for the 8 strata case, defines a set of eight half-sample replicates which will have the "cross-product balance" property. If it is necessary to use the eighth column, the resulting set of half samples will not have each element appearing an equal number of times. This will not destroy the variance estimating characteristics of the set of half samples, but it does mean that the average of the eight half-sample means will not necessarily be equal to the overall sample mean.

Since orthogonal matrices of plus and minus ones can be obtained whenever the order of the matrix is a multiple of four, it is always possible to find a set of half-sample replicates having cross-product balance. It follows that the number of half samples required will be at most three more than the number of strata. The HES analysis is based on 21 strata and it is therefore necessary to use 24 half-sample replicates. A possible set of balanced half samples for this particular case is shown below. This design was obtained by using the first 21 columns of the construction given in the Plackett-Burman paper. Any two columns of this design are orthogonal, and each element appears in 12 of the 24 replicates. The entire pattern is completely determined by the first column. The second column is obtained from the first by moving each sign down one position and placing the 23d sign at the top of the second column. This rotation is applied repeatedly to obtain the remaining columns. The 24th position is always minus and is not involved in the rotation.

The staff of NCHS has adapted the foregoing method of balanced half-sample replication to the routine analysis of data from HES, and this adaptation is described in the appendix. Professor Leslie Kish of the Survey Research Center of the University of Michigan has also employed the method in estimating the sampling errors of multiple regression coefficients, although the results of his study are not yet available for publication.

Partially Balanced Half Samples

If a complex sample design is based on a large number of strata, say 80, then it would ordinarily be undesirable to use a complete set of balanced half samples. As we have seen, a complete set would consist of 80 samples, and the processing of results for such a large number of half samples might prove too costly. The question can then be raised of whether it is possible to devise a set of k partially balanced half samples which will produce a more precise estimate of variance than can be obtained from k independent half-sample replicates. One way of accomplishing this will now be described and evaluated.

To illustrate the approach, consider the case of $L = 4$. There are $2^4 = 16$ possible half samples; a balanced set of four half samples can be constructed as shown in the preceding section; we desire a partially balanced set of two half samples. The following design was proposed by the staff of NCHS:

Sample	*Stratum* 1	2	3	4
1	+	+	+	+
2	+	-	+	-

A plus indicates y_{h1} and a minus y_{h2}. As before, $d_h = (y_{h1} - y_{h2})$. This design was obtained by using orthogonal columns for the first two strata and then repeating this in the second two strata. The corresponding squared deviations of the half-sample means from the overall sample mean are:

$$(\bar{y}_{hs,1} - \bar{y}_{st})^2 = (1/4)\, \Sigma\, W_h^2 d_h^2 + (1/2) \sum_{h<k} W_h W_k\, d_h d_k$$

$$(\bar{y}_{hs,2} - \bar{y}_{st})^2 = (1/4)\, \Sigma\, W_h^2 d_h^2 + (1/2)(-W_1 W_2\, d_1 d_2 + W_1 W_3\, d_1 d_3 - W_1 W_4\, d_1 d_4 - W_2 W_3\, d_2 d_3 + W_2 W_4\, d_2 d_4 - W_3 W_4\, d_3 d_4)$$

Their sum, divided by two, is

$$(1/4) \sum W_h^2 d_h^2 + (1/2)(W_1 W_3\, d_1 d_3 + W_2 W_4\, d_2 d_4)$$

Thus we see that cross-product terms within orthogonal sets have been eliminated, and that a portion of the cross-product terms between orthogonal sets have been eliminated. Applying the

A set of balanced half samples for 21 strata[1]

Half sample	Stratum 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	+	-	-	-	-	+	-	+	-	-	+	+	-	-	+	+	-	+	-	+	+
2	+	+	-	-	-	-	+	-	+	-	-	+	+	-	-	+	+	-	+	-	+
3	+	+	+	-	-	-	-	+	-	+	-	-	+	+	-	-	+	+	-	+	-
4	+	+	+	+	-	-	-	-	+	-	+	-	-	+	+	-	-	+	+	-	+
5	+	+	+	+	+	-	-	-	-	+	-	+	-	-	+	+	-	-	+	+	-
6	-	+	+	+	+	+	-	-	-	-	+	-	+	-	-	+	+	-	-	+	+
7	+	-	+	+	+	+	+	-	-	-	-	+	-	+	-	-	+	+	-	-	+
8	-	+	-	+	+	+	+	+	-	-	-	-	+	-	+	-	-	+	+	-	-
9	+	-	+	-	+	+	+	+	+	-	-	-	-	+	-	+	-	-	+	+	-
10	+	+	-	+	-	+	+	+	+	+	-	-	-	-	+	-	+	-	-	+	+
11	-	+	+	-	+	-	+	+	+	+	+	-	-	-	-	+	-	+	-	-	+
12	-	-	+	+	-	+	-	+	+	+	+	+	-	-	-	-	+	-	+	-	-
13	+	-	-	+	+	-	+	-	+	+	+	+	+	-	-	-	-	+	-	+	-
14	+	+	-	-	+	+	-	+	-	+	+	+	+	+	-	-	-	-	+	-	+
15	-	+	+	-	-	+	+	-	+	-	+	+	+	+	+	-	-	-	-	+	-
16	-	-	+	+	-	-	+	+	-	+	-	+	+	+	+	+	-	-	-	-	+
17	+	-	-	+	+	-	-	+	+	-	+	-	+	+	+	+	+	-	-	-	-
18	-	+	-	-	+	+	-	-	+	+	-	+	-	+	+	+	+	+	-	-	-
19	+	-	+	-	-	+	+	-	-	+	+	-	+	-	+	+	+	+	+	-	-
20	-	+	-	+	-	-	+	+	-	-	+	+	-	+	-	+	+	+	+	+	-
21	-	-	+	-	+	-	-	+	+	-	-	+	+	-	+	-	+	+	+	+	+
22	-	-	-	+	-	+	-	-	+	+	-	-	+	+	-	+	-	+	+	+	+
23	-	-	-	-	+	-	+	-	-	+	+	-	-	+	+	-	+	-	+	+	+
24	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

[1]It is assumed that two elements are drawn independently from each of 21 strata. A plus sign is arbitrarily assigned to one of the elements and a minus sign to the other. A row in this table then identifies a particular sample, the element with the indicated sign being taken from each of the 21 strata.

argument used in obtaining equation (4.8), the rel-variance of the estimated variance is approximated by

$$\frac{2 \times 2}{2 \times 4} + \frac{\beta + 1}{2 \times 4} = \frac{1}{2} + \frac{\beta + 1}{8}$$

If the two half-sample replicates had been chosen independently this rel-variance would be, from equation (4.8),

$$\frac{3}{4} + \frac{\beta+1}{8}$$

Hence a gain has been achieved, over the use of independent replicates, by using a partially balanced set of half samples.

This argument can be set forth in general terms. Suppose there are L strata and that a balanced set of k half samples is used, where for convenience we assume that L divided by k is an integer, and the k-pattern of half samples is repeated over each succeeding set of k strata. (In the preceding example, $L = 4$, $k = 2$, and $L/k = 2$.) Then

$$\frac{\sum_{i=1}^{k} (\bar{y}_{hs,i} - \bar{y}_{st})^2}{k} = (1/4) \sum W_h^2 d_h^2 + (1/2) \sum_{h,j} W_h W_j d_h d_j \qquad (4.9)$$

where the second summation is taken over all pairs (h, j) such that

$h < j$

h is from one set of k strata and j is from a second set of k strata

h and j represent corresponding columns from the k orthogonal columns that make up a balanced set.

The remaining cross-product terms disappear because of orthogonality conditions. It is easy to see that there are

$$k_{(L/k)}C_2 = \frac{k \times \frac{L}{k}(\frac{L}{k} - 1)}{2} = \frac{L(L-k)}{2k}$$

terms in the summation.

Since h and j are always different in the second part of equation (4.9) and since $E(d_h) = 0$, the covariance of the two parts of (4.9) is equal to zero. Therefore the variance of the expression is the sum of the separate variances. If one now assumes that the W_h's are equal and that the S_h^2's are all equal, the argument that led to expression (4.8) gives for the rel-variance of the estimated variance

$$\frac{2(L-k)}{kL} + \frac{\beta + 1}{2L} \qquad (4.10)$$

Since the first term in expression (4.8) is $2(L-1)/kL$, we see that the reduction in this portion of the rel-variance is like a finite population correction effect. By way of illustration, let us take $L = 80$, $k = 20$ or 40, and $\beta = 3$ or 6. The results are given in table 4. The gains for 20 replications are modest, but they are essentially achieved at little or no expense.

In conclusion, we observe that Gurney (1964) has approached this problem in a different fashion. Suppose again that there are 80 strata and that one enumerates the complete set of 80 balanced half samples. If one now selects without replacement 20 replicates from the complete set of 80, Gurney shows that the gains over independent replication are essentially the same as for the procedure that has just been considered. Thus two alternative ways of proceeding are available, both giving finite population factor gains in one component of the total variance.

Half-Sample Replication and the Sign Test

Suppose that the sample observations y_{h1} and y_{h2} with which we have been dealing actually represent the differences of two other variables, say z_{hi} and x_{hi}. That is, $y_{h1} = (z_{h1} - x_{h1})$ and $\bar{y}_{h2} = (z_{h2} - x_{h2})$. Under these circumstances, the population mean of the variable y is equal to the difference of the population means of the variables z and x. If the assumption of approximate normality for $\bar{y}_{st}$ is questioned, or if one wishes to avoid the computation of variances, it is tempting to think of combining the half-sample replication technique and a nonparametric technique in the following manner.

Obtain k half-sample estimates of $\bar{Y}$. As before, denote these by $\bar{y}_{hs,1}, \bar{y}_{hs,2}, \ldots, \bar{y}_{hs,k}$.

Apply some type of nonparametric technique to these observations. For example, one might use a sign test to test hypotheses about the median of the distribution from which the $\bar{y}_{hs,i}$ are drawn, or one might use the order statistics of this set of

Table 4. Comparison of components of rel-variance for independent replicates and for partially balanced replicates[1]

(L = 80)

		$\beta = 3$	$\beta = 6$
$k = 20$	Independent replicates	.099 + .025 = .124	.099 + .044 = .143
	Partially balanced replicates	.075 + .025 = .100	.075 + .044 = .119
		(.100)/(.124) = 81%	(.119)/(.143) = 83%
$k = 40$	Independent replicates	.049 + .025 = .074	.049 + .044 = .093
	Partially balanced replicates	.025 + .025 = .050	.025 + .044 = .069
		(.050)/(.074) = 68%	(.069)/(.093) = 74%

[1]The first term in each sum represents the component due to half-sample replication. The second term represents the component that arises in ordinary variance estimation.

$\bar{y}_{hs,i}$'s to determine confidence intervals for a population median.

The difficulties associated with such an analysis arise from the fact that the half-sample replicates are not independent of one another, when they are viewed as samples from the overall population. As an illustration, let us consider the following two half-sample estimates:

$$\bar{y}_{hs,1} = W_1 y_{11} + W_2 y_{21} + W_3 y_{31}$$

$$\bar{y}_{hs,2} = W_1 y_{11} + W_2 y_{22} + W_3 y_{32}$$

If repeated drawings of the entire sample are made, then

$$E(\bar{y}_{hs,i}) = W_1 \bar{Y}_1 + W_2 \bar{Y}_2 + W_3 \bar{Y}_3$$

$$V(\bar{y}_{hs,i}) = W_1^2 S_1^2 + W_2^2 S_2^2 + W_3^2 S_3^2$$

$$\text{Cov}(\bar{y}_{hs,1}, \bar{y}_{hs,2}) = E\Big\{\big[W_1(y_{11} - \bar{Y}_1) + W_2(y_{21} - \bar{Y}_2) + W_3(y_{31} - \bar{Y}_3)\big] \times \big[W_1(y_{11} - \bar{Y}_1) + W_2(y_{22} - \bar{Y}_2) + W_3(y_{32} - \bar{Y}_3)\big]\Big\}$$

$$= W_1^2 S_1^2$$

All other cross-product terms in the covariance drop out because of the independence of selections within and between strata. Finally, the correlation between $\bar{y}_{hs,1}$ and $\bar{y}_{hs,2}$ is

$$\rho_{\bar{y}_{hs,1}, \bar{y}_{hs,2}} = \frac{W_1^2 S_1^2}{W_1^2 S_1^2 + W_2^2 S_2^2 + W_3^2 S_3^2}$$

If $W_1^2 S_1^2 = W_2^2 S_2^2 = W_3^2 S_3^2$, then

$$\rho_{\bar{y}_{hs,1}, \bar{y}_{hs,2}} = 1/3$$

and we see, in general, that the correlation between any two half-sample means will be equal to the fraction of common elements, provided that the $W_h S_h$ are all equal. Thus ρ can take on the values $0, 1/L, 2/L, 3/L, \ldots, (L-1)/L, 1$.

If one makes independent selections of half-sample replicates, then the number of elements common to any two replicates will be a random variable whose average will be $L/2$, and there is little one can do in a precise way about evaluating the effect of the correlation on a sign test. However, the completely balanced set of replicates that was described in "Balanced Half-Sample Replication" does have "nice" properties in this respect. For example, the set of four replicates given in this section for $L = 3$ is such that any pair of replicates have a single common element. Therefore the correlation between any two of these $\bar{y}_{hs}$'s is 1/3, assuming equality of the $W_h^2 S_h^2$.

Furthermore, these $\bar{y}_{hs}$'s have the same expected value and variance.

If one were to apply a sign test to these four $\bar{y}_{hs}$'s for the purpose of testing the hypothesis that median of the distribution from which they are drawn is zero, then one would look at the number of positive $\bar{y}_{hs}$'s in the sample of four. Ordinary theory assumes that the observations are independent, and that the probability that any single observation is greater than zero is 1/2. Hence, if the hypothesis is rejected whenever there are either zero or four positive observations, the level of significance is $(1/2)^4 + (1/2)^4 = .125$. This value does not hold in the present situation because observations are dependent upon one another.

It is possible, however, to obtain an appropriate value in this case if one is willing to assume that the $\bar{y}_{hs}$'s have a normal multivariate distribution. Gupta (1963, p. 817) provides a table which gives the "probability that N standard normal random variables with common correlation ρ are simultaneously less than or equal to H." If we take $N = 4, \rho = 1/2$, and $H = 0$, we find this probability to be .14974. Therefore the level of significance of the test procedure is 2(.14974)= .299 instead of the .125 that would be obtained from independent observations. This is a clear indication that the nominal significance levels of a sign test will be changed markedly if there is correlation among the observations.

Whether or not the analysis of the preceding paragraph can be generalized to a larger number of strata depends upon the pattern of common elements in the balanced half-sample replicates and upon the availability of tables. The situation regarding common elements can be inferred from an examination of the 8 x 8 orthogonal matrix given in "Balanced Half-Sample Replication." Because the matrix is orthogonal, any pair of rows of the full 8 x 8 matrix will contain four elements in common and the common correlation between half-sample means will be $\rho = 1/2$, assuming that the $W_h S_h$ are constant. This will hold in general if the number of strata is a multiple of four. If the number of strata, L, is one less than a multiple of four, and if one deletes the column whose sign entries are all the same then any pair of rows will still have the same number of elements in common. This will be $\frac{L+1}{2} - 1$, and the common correlation will be

$$\rho = \frac{\frac{L+1}{2} - 1}{L} = \frac{L-1}{2L}$$

For $L = 7$ the number of required half-sample replicates will be 8, the number of common elements between pairs of replicates will be 3, and $\rho = 6/14 = .43$. Using the Gupta tables for $N = 8$ and $\rho = .40$, the probability of obtaining all plus or all minus signs is 2 x (.07909)= .158. The corresponding value for eight independent observations is $(1/2)^8 + (1/2)^8 = .0078$. Again we see the marked disparity between the two significance levels.

If the number of strata is two less than a multiple of four, pairs of half samples will no longer have the same number of elements in common. It is, however, easy to show the following

Number of common elements	*Number of pairs of samples*
$\frac{L+2}{2} - 1$	$(L+2)(L+2)/4$
$\frac{L+2}{2} - 2$	$\frac{L(L+2)/4}{(L+1)(L+2)/2}$

One might then consider using an average correlation coefficient. A similar analysis could be used for cases where the number of strata is three less than a multiple of four, and one would then find three possible values for the "number of common elements."

At the present time, this type of analysis is not very practical because Gupta's tables give probabilities only for the case where *all* values are positive or *all* are negative. For larger numbers of strata, one would have to be able to move in from the extremes on the distribution of number of plus signs in order to get meaningful levels of significance. Lacking such tables, the preceding analysis shows that levels of significance are "too far off" when obtained from the assumption of independence. Actually, if one must assume normality in order to obtain probabilities for making a sign test, then one might just as well use normal tests for correlated observations, as described by Walsh (1947).

Jackknife Estimates of Variance From Stratified Samples

Quenouille (1956) introduced a method for adjusting the members of a certain class of biased estimators in order to reduce the bias from order $1/n$ to order $1/n^2$. Numerous other authors—Jones (1956, 1965), Tukey (1958), Durbin (1959), Deming (1963), Lauh and Williams (1963), Robson and Whitlock (1964), Miller (1964), and Brillinger (1964)—have made contributions to the development and behavior of such estimators and to the estimation of their variances. Tukey (1958) gives, in an abstract, the following brief account of these characteristics:

> The linear combination of estimates based on all the data with estimates based on parts thereof seems to have been first treated in print as a means of reducing bias by Jones (*J. Am. statist. Ass.*, Vol. 51 (1956), pp. 54-83). Let $y_{(.)}$ be the estimate based on all the data, $y_{(i)}$ that based on all but the ith piece, $\bar{y}_{(i)}$ the average of the $y_{(i)}$. Quenouille (*Biometrika*, Vol. 43 (1956), pp. 353-360) has pointed out some of the advantages of $ny_{(.)} - (n-1)\bar{y}_{(i)}$ as such an estimate of much reduced bias. Actually, the individual expressions $ny_{(.)} - (n-1)y_{(i)}$ may, to a good approximation, be treated as though they were independent estimates. Not only is each nearly unbiased, but their average sum of squares of deviations is nearly $n(n-1)$ times the variance of their mean, etc. In a wide class of situations they behave rather like projections from a non-linear situation on to a tangent linear situation. They may thus be used in connection with standard confidence procedures to set closely approximate confidence limits on the estimate.

This general procedure, as decribed by Tukey, has been called the Jackknife method since "Like a boy scout jackknife, such a technique should be usable for anything... although, again like a jackknife, many of its jobs could be better done by the corresponding specialized tool, if that tool were only at hand."

The application of these ideas to stratified sampling, with two independent selections per stratum, is straightforward. Consider the simple stratified model described in "Half-Sample Replication Estimates of Variance From Stratified Samples." For simplicity, let $L = 3$. Then if one leaves out each "piece" in succession, and forms the estimate based on the remaining "pieces," we have the following six estimates of the population mean:

$$q_{(1)} = W_1 y_{11} + W_2(y_{21} + y_{22})/2 + W_3(y_{31} + y_{32})/2$$

$$q_{(2)} = W_1 y_{12} + W_2(y_{21} + y_{22})/2 + W_3(y_{31} + y_{32})/2$$

$$q_{(3)} = W_1(y_{11} + y_{12})/2 + W_2 y_{21} + W_3(y_{31} + y_{32})/2$$

$$q_{(4)} = W_1(y_{11} + y_{12})/2 + W_2 y_{22} + W_3(y_{31} + y_{32})/2$$

$$q_{(5)} = W_1(y_{11} + y_{12})/2 + W_2(y_{21} + y_{22})/2 + W_3 y_{31}$$

$$q_{(6)} = W_1(y_{11} + y_{12})/2 + W_2(y_{21} + y_{22})/2 + W_3 y_{32}$$

In general, there will be $2L$ such quantities. The estimation procedure applied to the entire sample gives

$$q = W_1(y_{11} + y_{12})/2 + W_2(y_{21} + y_{22})/2 + W_3(y_{31} + y_{32})/2$$

If the original sample had consisted of six independent observations, then the Jackknife procedure would call for obtaining six estimates of the form $q^*_{(i)} = 6q - 5q_{(i)}$. In the present example, however, there are two independent selections within each stratum. Hence it is appropriate to form the six estimates

$$q^*_{(i)} = 2q - q_{(i)}$$

The final estimate of the population mean is

$$q^* = \sum_{i=1}^{6} q^*_{(i)}/6$$

and the variance of q^* is estimated by

$$\sum_{i=1}^{6} (q^*_{(i)} - q^*)^2/2$$

In general, the mean will be estimated by

$$q^* = \sum_{i=1}^{2L} q^*_{(i)}/2L$$

and its variance by

$$\sum_{i=1}^{2L} (q^*_{(i)} - q^*)^2/2$$

In the simple linear case that is being considered here, it is easy to show that an analysis in terms of the $q^*_{(i)}$'s produces results that are identical with those obtained by the standard analysis. For example,

$$q_{(1)} = \bar{y}_{st} + W_1 d_1/2, \quad q^*_{(1)} = \bar{y}_{st} - W_1 d_1/2$$

$$q_{(2)} = \bar{y}_{st} - W_1 d_1/2, \quad q^*_{(2)} = \bar{y}_{st} + W_1 d_1/2$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$q_{(5)} = \bar{y}_{st} + W_3 d_3/2, \quad q^*_{(5)} = \bar{y}_{st} - W_3 d_3/2$$

$$q_{(6)} = \bar{y}_{st} - W_3 d_3/2, \quad q^*_{(6)} = \bar{y}_{st} + W_3 d_3/2$$

$$q = \bar{y}_{st} \qquad q^* = \bar{y}_{st}$$

Furthermore

$$\sum_{i=1}^{6} (q^*_{(i)} - q^*)^2/2 = (W_1^2 d_1^2 + W_2^2 d_2^2 + W_3^2 d_3^2)/4$$

which is the ordinary estimate of the variance of $\bar{y}_{st}$, just as in the case of the balanced half-sample replicates.

There is another variant of the Quenouille type of estimate which is closely related to the half-sample approach. Suppose that a particular half sample is chosen. Denote its elements by $y_{11}, y_{21}, \ldots, y_{L1}$ and its mean by $\bar{y}_{hs}$. The set of remaining elements, one in each stratum, then constitutes an independent half sample whose mean we denote by $\bar{y}^*_{hs}$. A Quenouille-type estimate is then defined by

$$2\bar{y}_{st} - (\bar{y}_{hs} + \bar{y}^*_{hs})/2$$

which, in the simple linear case that is being considered here, is identically equal to $\bar{y}_{st}$. This approach does not provide an estimate of variance in the present instance since only one estimate is obtained. In more complicated situations, however, different half samples will provide different values of the Quenouille half-sample estimate, and it might be possible to base estimates of variance on these different values. This possibility will be discussed briefly in the following section.

Half-Sample Replication and the Jackknife Method With Stratified Ratio-Type Estimators

We have introduced half-sample replication and the Jackknife method in the setting of a simple linear situation, where they obviously have no real utility. Under these circumstances, they merely reproduce results that can be obtained by direct analysis. If, however, more complicated methods of sampling and estimation are employed, then direct methods of analysis may not be available, may require a prohibitive amount of computation in comparison with the methods being considered here, or may even give results that are in one way or another inferior to those provided by half-sample replication and the Jackknife.

Although one may accept on intuitive grounds the general premise that half-sample replication and the Jackknife do permit the "easy" computation of variance estimates that in one way or another mirror most of the standard complexities of sample design and estimation, the exact characteristics of the resulting estimates and their corresponding estimates of variance are, for the most part, unknown. This is particularly true for half-sample replication, even though the intuitive appeal of this method may be more direct than that of the Jackknife. No published or unpublished references to the behavior of half-sample replication in complex situations were discovered, and the notion of balanced half samples was introduced in this report for the first time. On the other hand, there is a growing body of literature and data relating to the Jackknife. We shall now summarize this material on the Jackknife and then report the results of a very small experiment which compares results obtained by balanced half-sample replication and by the Jackknife.

Although Quenouille (1956) introduced his method of adjustment as a means for reducing the bias of an estimator, our interest in the Jackknife is primarily focused on its utility for variance estimation. One is naturally interested in obtaining any reductions in bias that are possible, but there is a considerable body of empirical evidence—notably in the work of Kish, Namboodiri, and Pillai (1962)—which indicates that the "combined ratio estimator" for population means, subpopulation means, and differences of subpopulation means probably has negligible bias in most practical surveys. On p. 863, Kish et al. say "Our empirical investigations, set in a theoretical framework, show that the bias in most practical surveys is usually negligible; the ratio of bias to standard error (B/σ) was small in every test, even those based on small sub-classes."

There is actually very little published material which has a direct bearing on our present concern. The pertinent items are briefly summarized.

1. Quenouille (1956) has shown by formal analysis that the variance of his estimator, where such an estimator is appropriate, differs from the variance of the unadjusted estimator by terms of order $1/n^2$.
2. Durbin (1959) applied the method of Quenouille to the ratio estimator $r = \bar{y}/\bar{x}$, where a random sample was divided into two groups of equal size. Thus he considered the estimator, of $E(y)/E(x)$,

$$t = 2r - (r_1 + r_2)/2$$

   where $r = (y/x)$, $r_1 = (y_1/x_1)$, $r_2 = (y_2/x_2)$, y and x are sample totals, y_1 and x_1 are half-sample totals, and y_2 and x_2 are the other half-sample totals. Durbin considers two cases: (1) x is a normal variable with variance $O(n^{-1})$, and the regression of y on x is linear, not necessarily through the origin; (2) x is a gamma variable with mean m and the regression of y on x is linear. For the first case, when terms of $O(n^{-4})$ are ignored, the result is obtained that the variance of t is smaller than the variance of r. For the second case, it was not necessary to use an approximate form of analysis. Durbin concludes that ". . . whenever the coefficient of variation of x is less than 1/4, which will be satisfied by all except the most inaccurate estimators, Quenouille's estimator has a smaller mean square error than the ordinary ratio estimator. This is an exact result for any sample size."
3. Brillinger (1964) studies the properties of these estimators in relation to maximum likelihood estimators. His conclusions are: "Summing up the results of the paper, one may say that Tukey's general technique of setting approximate confidence limits is asymptotically correct, under regularity conditions, when applied to maximum likelihood estimates and that the technique provides a useful method of estimating the variance of an estimate. Also one may say that the estimate proposed by Quenouille will on many occasions have reduced variance, smaller mean-squared error and closer to asymptotic normality properties, when compared to the usual maximum likelihood estimates."
4. Robson and Whitlock (1964) apply Quenouille's method of construction to obtain estimates of a truncation point of a distribution. One of the interesting features of their work relates to the construction of estimators that successively eliminate bias terms of order n^{-2}, n^{-3}, etc. They find for their particular problem that the variance of these estimators increases as the bias is decreased.
5. Miller (1964) is concerned with conditions under which a Jackknife estimator, and its

associated estimated variance, will asymptotically have a Student's t distribution. Both of the situations described by Miller are ones in which the unjackknifed estimator had a proper finite or limiting distribution under weaker conditions than required for the Jackknife.

The foregoing five references are concerned with estimators and with their bias and variance. None deals with the problem of estimating variances. However, Lauh and Williams (1963) do present some Monte Carlo results which relate primarily to the estimation of variance. They are again concerned with the estimation of a ratio, $E(y)/E(x)$, but the Quenouille procedure is applied to the individual sample observations instead of to half samples as in the Durbin investigation. That is, they compare the behavior of

$$q = (\sum_{i=1}^{n} y_i)/(\sum_{i=1}^{n} x_i)$$

with the behavior of

$$q^* = \sum_{i=1}^{n} q^*_{(i)}/n$$

where

$$q_{(i)} = (\sum_{j=1}^{n} y_j - y_i)/(\sum_{j=1}^{n} x_j - x_i)$$

and

$$q^*_{(i)} = nq - (n-1)\, q_{(i)}$$

This is similar to the estimator that was proposed for stratified sampling in the preceding section. Lauh and Williams define two populations which are used for empirical sampling: (1) x is a normal variable, while the regression of y on x is linear *through the origin;* (2) x is a chi-square variable with 2 degrees of freedom, and the regression of y on x is linear *through the origin.* Since the regressions are forced to go through the origin, both q and q^* are unbiased estimators of $E(y)/E(x)$. For each population 1,000 samples of n are drawn, $n = 2, 3, \ldots, 9$, and a variety of variance estimators are considered. In particular, the ordinary estimate of variance obtained from a Taylor series approximation was employed, denoted by $v_1(q)$; also an estimate of variance was obtained from the $q^*_{(i)}$'s, namely

$$v(q^*) = \sum_{i=1}^{n} (q^*_{(i)} - q^*)^2/n(n-1)$$

The results of this investigation are most interesting, particularly the fact that the precision of $v(q^*)$ is much better than that of $v_1(q)$ when x has an exponential distribution, and are summarized by the authors as follows:

> From the results of these two studies, it may be inferred that the bias of the estimator $v_1(q)$ is dependent upon the degree of skewness of the original y and x populations. Estimators of the true variance taken from higher order approximations lead only to slight improvements over the second order approximation $v_1(q)$, and in some cases the estimate is actually worse. The precision of $v(q^*)$ is nearly double that of $v_1(q)$ for exponential x distributions and the bias of $v(q^*)$ is smaller than that of $v_1(q)$. Thus it appears that the split-sample estimator q^* may be definitely preferable to q in some situations.

Finally, we note that extensive Monte Carlo investigations of many of these points have been initiated by Dr. Benjamin Tepping, Director of the Center for Measurement Research in the U.S. Bureau of the Census. Results of these investigations are not yet available.

The foregoing references are concerned only with random samples drawn from infinite populations. Our principal concern is with stratified samples drawn, usually without replacement, from finite populations where complex estimation procedures are applied to the basic sample data. Under these circumstances, a careful investigation of the behavior of estimators and of variance estimators would undoubtedly require a large-scale, Monte Carlo type of program, integrated with as much analytic work as possible. This was not feasible within the confines of the present study, even in terms of planning. Nevertheless, we did

Table 5. Artificial population

Stratum					
I		II		III	
y	x	y	x	y	x
3	4	5	4	7	3
4	6	9	8	9	4
11	20	24	23	25	12
Total 18	30	38	35	41	19
$R_1 = \frac{18}{30} = .6000$		$R_2 = \frac{38}{35} = 1.0857$		$R_3 = \frac{41}{19} = 2.1579$	
		$R = \frac{97}{84} = 1.1548$			

desire some small numerical model that would illustrate the various points that have been raised.

As an example, we started with the small artificial population that is used for illustrative purposes in Cochran (1963, p. 178, 179). However, since we wished to enumerate all possible samples and thereby investigate the behavior of both balanced half-sample replication and the Jackknife method of variance estimation, and since computations were to be carried out on a desk calculator, we were not able to use the full population as given by Cochran. (It did not appear worthwhile to invest computer programming time on one isolated and artificial example.) Accordingly one observation was dropped from each stratum and the following population was used as shown in table 5.

For this population, all possible samples of six, $n_h = 2$ in each stratum, were enumerated $- 3^3 = 27$ possible samples. For each sample, the following quantities were computed.

1. The combined ratio estimate.
2. The estimate of variance based on the ordinary Taylor series approximation to the variance of the combined ratio estimate.
3. The Quenouille estimate of the population ratio, using individual observations as previously described.
4. The Jackknife estimate of variance, as described in the preceding section.
5. The average of four balanced half-sample estimates, as described in "Balanced Half-Sample Replication." It was necessary to consider two sets of balanced half samples, one complementary to the other, since this is a nonlinear situation. It is assumed that one of these two sets will be chosen randomly in practical applications.
6. The estimate of variance based on the four balanced half-sample estimates. This estimate of variance is the sum of squared deviations of four half-sample estimates *about the combined ratio estimate*, divided by four, and multiplied by the finite population correction. This is the manner in which the half-sample estimate has been applied in the work of the U.S. Bureau of the Census and in HES. Again the estimate was made for each of the two sets of balanced half samples.
7. The Quenouille estimate based on the balanced half samples. That is, a Quenouille estimate was obtained from a half sample and its complement. This was carried out for each half sample and then averaged over the set of four balanced half samples. The results of these computations are summarized in table 6.

Table 6. Behavior of estimates and estimates of variance obtained by enumerating samples drawn from artificial population of table 5

Estimate	Bias	Standard error	Variance	Mean square error	Average variance estimate	Variance of the variance estimates
Combined ratio estimate------	+.0118	.122	.0148	.0149	.0099	.000034
Quenouille estimate, individual observations-----	+.0034	.126	.0160	.0160	.0110	.000040
Balanced half-sample estimate-------------------	+.0428	.104	.0109	.0127	.0122	.000047
Quenouille estimate, half samples----------------	-.0198	.137	.0188	.0192	...	...

In obtaining the variance estimates a finite population correction of $(1-\frac{2}{3})$ was applied uniformly. It can be readily demonstrated that this is appropriate for the Jackknife and balanced half-sample estimates of variance, at least in the simple linear case that was used to introduce these techniques. Jones (1965) describes a modification of the Jackknife, whose purpose is to introduce the finite population correction into the "bias-reducing" argument upon which the Quenouille adjustment rests. This modification was not used here.

Although it is clearly impossible to draw any general conclusions from one artificial example such as the above, the results are interesting. In particular, we find that the combined ratio estimate has almost negligible bias and that the ordinary variance estimate seriously underestimates the true variance; that the Quenouille estimate with individual observations does reduce the bias at the expense of increasing the variance by about 8 percent, and that the Jackknife estimate of variance again seriously underestimates the true variance; that the average of the four balanced half-sample estimates has the smallest variance and the largest bias, while at the same time providing a reasonable estimate of the corresponding variance. The variance of the variance estimates is largest for the estimate based on the balanced half samples. Even if one makes the comparison on the basis of mean square error, the balanced half-sample average is still superior to the other estimates in spite of its larger bias, except for the variance of the variance estimate. As a final point, the Quenouille adjustment applied to the balanced half samples does reduce the bias, but at the expense of a marked increase in terms of variance.

This example, trivial and artificial as it may be, does raise one question that concerns half-sample replication. The use of the Quenouille estimate and the Jackknife method of estimating variances have usually been considered simultaneously. On the other hand, the half-sample replication method of estimating variances has always been used to estimate the variance of the estimate obtained from the entire sample, and not the variance of the average of the half-sample estimates. Of course when one does not have an "exhaustive" set of half samples, as in the case where they are drawn with replacement, the average of half-sample estimates would not be appropriate. Here, however, we do have an "exhaustive" set of balanced half-sample estimates, and we might well consider using their average in place of the combined ratio estimate. In terms of our example, a variance estimate with average value .0122 is being used to estimate the variance of the combined ratio estimate, whose true value is .0148. This is somewhat better than the ordinary Taylor series variance estimate, whose average value is .0099, but not nearly as good as if one uses the half-sample estimate of variance to estimate the mean square error of the average of the balanced half-sample estimates—namely, a quantity whose average is .0122 to estimate a quantity whose true value is .0127.

A small amount of data which relates to the foregoing point is presented in table IV of the appendix. For each of the six subclasses for which comparisons of percentages are presented, the estimate obtained from the entire sample can be compared with the average of the 16 balanced half-sample estimates. It follows from the argument used in developing the Quenouille-type estimate that the difference between the two can be viewed as an estimate of bias for the overall estimate. This approach has been used by Deming (1960, p. 425). These estimates of bias, expressed as fractions of the estimated standard errors, range from approximately .03 to about .38. These data are reassuring, but they are also too fragmentary to support any general conclusions concerning the bias of estimates for HES analysis.

In conclusion, attention should be drawn to another avenue of approach to estimation and variance estimation which is somewhat related to the Jackknife method, although this relation has not been explored or even noticed in the literature. Mickey (1959) presents a general method for obtaining finite population unbiased ratio and regression estimators building on the work of Goodman and Hartley (1958). In addition, he constructs unbiased estimates of the estimator variance by the process of breaking up the sample into subsamples, more or less along the lines of Tukey's general version of the Jackknife. Williams (1958, 1961) specializes these results to regression estimators, and considers their properties in some detail. No detailed attention has been given to this topic in connection with the preparation of the present report.

SUMMARY

Sampling theory provides a wide variety of techniques which can be applied in sample design to obtain estimates having essentially maximum precision for fixed cost. These techniques are particularly useful when populations are spread over wide geographic areas so that highly clustered samples must be obtained, and when extensive prior information about the population under study can be used in sample selection or in estimation. Such complex sample designs do, however, require extremely complicated and only approximate expressions for estimating from a sample the variance of survey estimates. If an extremely large number of widely differing types of estimates are to be made from a single large-scale survey, the burden of developing appropriate variance expressions, of programming these for a computer, and of carrying out the computations may become excessive.

The foregoing problems are intensified, although not appreciably changed in kind, if survey analysis is to go much beyond the estimation of population means, percentages, and totals. This is particularly true when the goals of analysis are to compare and study the relationships among subpopulations, or domains of study. Investigators are then interested in applying such standard statistical techniques as multiple regression or analysis of variance, and find that many of the assumptions required for the application of these techniques are violated by the complexities of the sample design. Some authors have used the term "analytical survey" to refer to any survey in which extensive comparisons are made among subpopulations; other authors reserve this term for surveys that are specifically designed to control the precision of these comparisons. There seemed to be little point in arguing this issue in the present report, since most surveys are multipurpose in character and it is usually impossible to design for a specific comparison. The major portions of the first three sections of the report are devoted to a literature survey and discussion of these topics.

Survey design (as opposed to the analysis of survey data) requires the use of "exact" variance expressions since it is necessary to balance the effects on precision of a wide variety of sampling techniques. It is possible, however, to bypass the corresponding detailed variance estimation techniques in the actual analysis of survey data through the use of replication. This approach is discussed in "Replication Methods of Estimating Variances," where an attempt is made to set forth its advantages and limitations. Emphasis has been placed upon variance estimation, although it is clear that covariances can also be treated in the same manner.

One of the most serious limitations of replication as applied to the analysis of complex sample survey data arises from the difficulty of obtaining a sufficient number of independent replications to assure reasonably stable variance estimates. This fact has been particularly obvious when a set of primary sampling units is stratified to a point where the sample design calls for the selection of two primary sampling units per stratum, thus leading to only two independent replicates. To overcome this difficulty the U.S. Bureau of the Census and the National Center for Health Statistics have been using a pseudoreplication method for variance estimation, called half-sample replication. This procedure is described in "Half-Sample Replication Estimates of Variances From Stratified Samples," and several improvements, balanced half-sample replication and partially balanced half-sample replication, are introduced in "Balanced Half-Sample Replication" and "Partially Balanced Half Samples." Still another, but related, variance estimation technique, the Jackknife, is described in "Jackknife Estimates of Variance From Stratified Samples." The application of these methods is illustrated on an artificial set of data in "Half-Sample Replication and the Jackknife Method With Stratified Ratio-Type Estimators," and the appendix shows how balanced half-sample replication has been used in analyzing data obtained from the Health Examination Survey.

It would appear that replication and pseudoreplication are extremely useful procedures for obtaining variance estimates when one is making detailed analyses of data derived from complex sample surveys. Nevertheless, there are many unresolved problems relating to the application of these methods. Among these are the following: (1) The effects of certain sampling techniques on variances will not be picked up—e.g., the selection of one primary unit per stratum and controlled selection; (2) The variance estimate ordinarily refers to the average of the replication estimates, whereas the ordinary procedure is to use an overall sample estimate, and the two will not be the same except in the rare case that the estimate is linear in form; (3) No investigations have been carried out of the applicability of these approaches to such problems as contingency table analyses and standard analysis of variance approaches; and (4) It is extremely difficult to attack any of these problems analytically, and the development of empirical approaches that will have widespread applicability seems most difficult.

As a final point, we call attention to the problems that arise when survey data, as opposed to experimental data, are used to develop general scientific conclusions. This topic has not been more than mentioned in this report, but reference may be made to discussions by Yates (1960), Kish (1959), and Blalock (1964).

BIBLIOGRAPHY

Aoyama, H.: A study of the stratified random sampling. *Annals of the Institute of Statistical Mathematics* VI(1):1-36, 1954-55.

Aspin, A.: Tables for use in comparisons whose accuracy involves two variances, separately estimated. *Biometrika* 36:290-293, 1949.

Bartlett, M. S.: The information available in small samples. *Proc.Camb.Phil.Soc.* 32:560, 1936.

Blalock, H. M., Jr.: *Causal Inferences in Nonexperimental Research.* Chapel Hill. University of North Carolina Press, 1964.

Brillinger, D. R.: The asymptotic behavior of Tukey's general method of setting confidence levels (the Jackknife) when applied to maximum likelihood estimates. *Review of International Statistical Institute* 32:202-206, 1964.

Cochran, W. G.: *Sampling Techniques,* ed. 2. New York. John Wiley and Sons, Inc., 1963.

Cornfield, J.: On samples from finite populations. *J.Am. statist.Ass.* 39(226):236-239, June 1944.

Cornfield, J., and Tukey, J. W.: Average values of mean squares in factorials. *Ann.math.Statist.* 27:907-949, 1956.

Deming, W. E.: *Some Theory of Sampling.* New York. John Wiley and Sons, Inc., 1950.

Deming, W. E.: *Sample Design in Business Research.* New York. John Wiley and Sons, Inc., 1960.

Deming, W. E.: On the correction of mathematical bias by use of replicated designs. *Metrika* 6:37-42, 1963.

Deming, W. E., and Stephan, F. F.: On the interpretation of censuses as samples. *J.Am.statist.Ass.* 36(213):45-49, Mar. 1941.

Durbin, J.: Sampling theory for estimates based on fewer individuals than the number selected. *Bulletin of the International Statistical Institute* 36:113-119, 1958.

Durbin, J.: A note on the application of Quenouille's method of bias reduction to the estimation of ratios. *Biometrika* 46:477-480, 1959.

Durbin, J., and Stuart, A.: Differences in response rates of experienced and inexperienced interviewers. *Journal of the Royal Statistical Society,* Series A (General), 114, Pt. II, pp. 163-206, 1951.

Fisher, R. A.: *The Design of Experiments,* ed. 3. London. Oliver and Boyd, Ltd., 1942.

Gautschi, W.: Some remarks on systematic sampling. *Ann. math.Statist.* 28:385-394, 1957.

Garza-Hernandez, T.: *An Approximate Test of Homogeneity on the Basis of a Stratified Random Sample.* M. S. thesis, New York State School of Industrial and Labor Relations, Cornell University, 1961.

Goodman, L. A., and Hartley, H. O.: The precision of unbiased ratio-type estimators. *J.Am.statist.Ass.* 53(282):491-508, June 1958.

Goodman, R., and Kish, L.: Controlled selection—a technique in probability sampling. *J.Am.statist.Ass.* 45(251):350-372, Sept. 1950.

Gupta, S. S.: Probability integrals of multivariate normal and multivariate *t*. *Ann.math.Statist.* 34:792-828, 1963.

Gurney, M.: *The Variance of the Replication Method for Estimating Variances for the CPS Sample Design.* Unpublished memorandum, U.S. Bureau of the Census, 1962.

Gurney, M.: *McCarthy's Orthogonal Replications for Estimating Variances, With Grouped Strata.* Unpublished memorandum, U.S. Bureau of the Census, 1964.

Hansen, M. H., Hurwitz, W. N., and Madow, W. G.: *Sample Survey Methods and Theory,* Vols. I and II. New York. John Wiley and Sons, Inc., 1953.

Hanson, R. H., and Marks, E. S.: Influence of the interviewer on the accuracy of survey results. *J.Am.statist.Ass.* 53(283):635-655, Sept. 1958.

Hartley, H. O.: *Analytic Studies of Survey Data.* Instituto di Statistica, Rome, Volume in onora di Corrado Gini. Ames, Iowa. Statistical Laboratory, Iowa State University of Science and Technology. Reprint Series 63. 1959.

Jones, H. L.: Investigating the properties of a sample mean by employing random subsample means. *J.Am.statist.Ass.* 51(273):54-83, Mar. 1956.

Jones, H. L.: The Jackknife method. *Proceedings of the IBM Scientific Computing Symposium on Statistics.* White Plains, New York. IBM Data Processing Division, 1965.

Keyfitz, N.: A factorial arrangement of comparisons of family size. *Am.J.Soc.* 58(5):470-480, Mar. 1953.

Keyfitz, N.: Estimates of sampling variance where two units are selected from each stratum. *J.Am.statist.Ass.* 52 (280):503-510, Dec. 1957.

Kish, L.: Confidence intervals for clustered samples. *American Sociological Review* 22(2):154-165, Apr. 1957.

Kish, L.: Some statistical problems in research design. *American Sociological Review* 24:328-338, June 1959.

Kish, L.: Efficient allocation of a multi-purpose sample. *Econometrica* 29:363-385, 1961.

Kish, L.: Variances for indexes from complex samples. *Proceedings of the Social Statistics Section of the American Statistical Association,* 1962, pp. 190-199.

Kish, L.: *Survey Sampling.* New York. John Wiley and Sons, Inc., 1965.

Kish, L., and Hess, I.: On variances of ratios and their differences in multistage samples. *J.Am.statist.Ass.* 54(286): 416-446, June 1959.

Kish, L., Namboodiri, N. K., and Pillai, R. K.: The ratio bias in surveys. *J.Am.statist.Ass.* 57(300):863-876, Dec. 1962.

Koop, J. C.: On theoretical questions underlying the technique of replicated or interpenetrating samples. *Proceedings* *of the Social Statistics Section of the American Statistical Association,* 1960, pp. 196-205.

Lahiri, D. B.: Technical paper on some aspects of the development of the sample design, in P. C. Mahalanobis "Technical Paper No. 5 on the National Sample Survey." *Sankhya* 14:264-316, 1954.

Lauh, E., and Williams, W. H.: Some small sample results for the variance of a ratio. *Proceedings of the Social Statistics Section of the American Statistical Association,* 1963, pp. 273-283.

Madow, W. G., and Madow, L. H.: On the theory of systematic sampling. *Ann.math.Statist.*XV:1-24, 1944.

McCarthy, P. J.: Sampling considerations in the construction of price indexes with particular reference to the United States Consumer Price Index. U.S. Congress, Joint Economic Committee, *Government Price Statistics, Hearings*. Washington. U.S. Government Printing Office. Part 1, pp. 197-232, 1961.

McCarthy, P. J.: Stratified sampling and distribution-free confidence intervals for a median. Accepted for publication in *J.Am.statist.Ass.,* 1965.

Meier, P.: Variance of a weighted mean. *Biometrics* 9(1): 59-73, Mar. 1953.

Mickey, M. R.: Some finite population unbiased ratio and regression estimators. *J.Am.statist.Ass.* 54(287):594-612, Sept. 1959.

Miller, R. G., Jr.: A trustworthy Jackknife. *Ann.math.Statist.* 35(4):1594-1605, 1964.

National Center for Health Statistics: Cycle I of the Health Examination Survey, sample and response, United States, 1960-62. *Vital and Health Statistics*. PHS Pub. No. 1000-Series 11-No. 1. Public Health Service. Washington. U.S. Government Printing Office, Apr. 1964.

National Center for Health Statistics: Blood pressure of adults by age and sex, United States, 1960-62. *Vital and Health Statistics*. PHS Pub. No. 1000-Series 11-No. 4. Public Health Service. Washington. U.S. Government Printing Office, June 1964.

National Center for Health Statistics: Blood pressure of adults by race and area, United States, 1960-62. *Vital and Health Statistics*. PHS Pub. No. 1000-Series 11-No. 5. Public Health Service. Washington. U.S. Government Printing Office, July 1964.

National Center for Health Statistics: Plan and initial program of the Health Examination Survey. *Vital and Health Statistics*. PHS Pub. No. 1000-Series 1-No. 4. Public Health Service. Washington. U.S. Government Printing Office, July 1965.

National Health Survey: The statistical design of the Health Household-Interview Survey. *Health Statistics*. PHS Pub. No. 584-A2. Public Health Service. Washington. U.S. Government Printing Office, July 1958.

Okamoto, M.: Chi-square statistic based on the pooled frequencies of several observations. *Biometrika* 50:524-528, 1963.

Plackett, R. L., and Burman, J. P.: The design of optimum multifactorial experiments. *Biometrika* 33:305-325, 1943-46.

Quenouille, M. H.: Notes on bias in estimation. *Biometrika* 43:353-360, 1956.

Robson, D. S., and Whitlock, J. H.: Estimation of a truncation point. *Biometrika* 51:33-39, 1964.

Satterthwaite, F. E.: An approximate distribution of estimates of variance components. *Biometrics* 2:110-114, 1946.

Sedransk, J.: *Sample Size Determination in Analytical Surveys*. Ph.D. dissertation, Harvard University, 1964.

Sedransk, J.: *Analytical Surveys With Cluster Sampling*. Unpublished paper, Iowa State University, 1964a.

Sedransk, J.: *A Double Sampling Scheme for Analytical Surveys*. Unpublished paper, Iowa State University, 1964b.

Sukhatme, P. V.: *Sampling Theory of Surveys With Applications*. Ames, Iowa. Iowa State College Press, 1954.

Tukey, J. W.: Bias and confidence in not-quite large samples. Abstracted in *Ann.math.Statist.* 29:614, 1958.

U.S. Bureau of the Census: *The Current Population Survey, A Report on Methodology*. Technical Paper No. 7. Washington. U.S. Government Printing Office, 1963.

Walsh, J. E.: Concerning the effect of intraclass correlation on certain significance tests. *Ann.math.Statist.* 18(1): 88-96, 1947.

Walsh, J. E.: On the "information" lost by using a *t*-test when the population variance is known. *J.Am.statist.Ass.* 44(245):122-125, Mar. 1949.

Weibull, M.: The distributions of *t*- and *F*-statistics and of correlation and regression coefficients in stratified samples from normal populations with different means. *Skand.Aktuar-Tidskr.* 36(suppl. 1, 2):9-106, 1953.

Welch, B. L.: The generalization of "Student's" problem when several different population variances are involved. *Biometrika* 34:28-35, 1947.

Williams, W. H.: *Unbiased Regression Estimators and Their Efficiencies*. Ph.D. dissertation, Iowa State College, 1958.

Williams, W. H.: Generating unbiased ratio and regression estimators. *Biometrics* 17(2):267-274, June 1961.

Yates, F.: *Sampling Methods for Censuses and Surveys,* ed. 3. New York. Hafner Publishing Co., 1960.

———o o o———

APPENDIX

ESTIMATION OF RELIABILITY OF FINDINGS FROM THE FIRST CYCLE OF THE HEALTH EXAMINATION SURVEY

Survey Design

The sampling plan of the first cycle of the Health Examination Survey followed a highly stratified, multistage probability design in which a sample of the civilian, noninstitutional population of the conterminous United States, 18-79 years of age, was selected. In the first stage of this design, the 1,900 primary sampling units (PSU's), geographic units into which the United States was divided, were grouped into 42 strata. Here a PSU is either a standard metropolitan statistical area (SMSA) or one to three contiguous counties. By virtue of their size in population, the six largest SMSA's were considered to be separate strata and were included in the first-stage sample with certainty. As New York was about three times the size of other strata and Chicago twice the average size, New York was counted as three strata and Chicago as two, making a total of nine certainty strata. One PSU was selected from among the PSU's in each of the 33 noncertainty strata to complete the first-stage sample. Later stages resulted in the random selection of clusters of typically four persons from segments of households within the sample PSU's. The total sampling included some 7,700 persons in 29 different States.

All examination findings for sample persons are included in tabulations as weighted frequencies, the weight being a product of the reciprocal of the probability of selecting the individual, an adjustment for nonresponse cases, a stratified ratio adjustment of the first-stage sample to 1960 Census population controls within 6 region-density classes, and a poststratified ratio adjustment at the national level to independent population controls for the midsurvey period (October 1961) within 12 age-sex classes.

The sample design is such that each person has roughly the same probability of selection. However, there were sufficient deviations from that principle in the selection and through the technical adjustments to produce the following distribution of sample weights as required to inflate to U.S. civilian, noninstitutional population levels:

Weight class	*Class average*	*1-digit relative weight*	*Percent distribution of examined persons*
7,000-20,999	14,000	1	78.7
21,000-34,999	28,000	2	18.4
35,000-48,999	42,000	3	1.9
49,000-62,999	56,000	4	0.6
63,000-76,999	70,000	5	0.0
77,000-90,999	84,000	6	0.4

A more detailed description of the sampling plan and estimation procedures is included in *Vital and Health Statistics*, Series 11, No. 1, 1964: "Cycle I of the Health Examination Survey, Sample and Response."

Requirements of a Variance Estimation Technique

The Health Examination Survey is obviously complex in its sampling plan and estimation procedure. A method for estimating the reliability of findings is required which reflects both the losses from clustering sample cases at two stages and the gains from stratification, ratio estimation, and poststratification. Ideally, an appropriate method once programmed for an electronic computer can be used for a wide range of statistics with little or no modification to the program. This feature of adaptability is an important and special requirement in HES. The small staff of analysts in the Division of Health Examination Statistics typically works on only a few sections of the examination and laboratory results at a time. Consequently tabulation specifications and edited input for a sizable variety of report topics are not available until shortly prior to the need for estimates of sampling error. New tables of sampling error have been prepared for each of the 12 reports published to date and at least a dozen more will be prepared before the Cycle I publication series is completed.

Development of the Replication Technique

The method adopted for estimating variances in the Health Examination Survey is the half-sample replication technique. The method was developed at the U.S. Bureau of the Census prior to 1957 and has at times been given limited use in the estimation of the reliability of results from the Current Population Survey. A description of the half-sample replication technique, however, has not previously been published, although some references to the technique have appeared in the literature.

The half-sample replication technique is particularly well suited to the Health Examination Survey because the sample, although complex in design, is relatively small (7,000 cases) in sample size. Only a few minutes are required for a pass of all cases through the computer. This feature permitted the development of a variance estimation program which is an adjunct to the general computer tabulation program. Every data table comes out of the computer with a table of desired estimates of aggregates, means, or distributions together with a table identical in format but with the estimated variances instead of the estimated statistics. The computations required by the method are indeed simple and the internal storage requirements are well within the limitation of an IBM 1401-1410 computer system.

The variance estimates computed for the first few. reports of Cycle I findings were based on 20 random half-sample replications. A half sample was formed by randomly selecting one sample PSU from each of 16 pairs of sample PSU's, the sample representatives of 16 pairings of similar noncertainty strata, and 8 of 16 random groups of clusters of sample persons selected from the 9 certainty strata and the San Francisco SMSA, the largest sample PSU of the 33 noncertainty strata. The concept of balanced half samples is utilized in present variance estimates for HES. The variance estimates are derived from 16 balanced half-sample replications. The composition of the 16 half samples, shown in tables I and II, was determined by an orthogonal plan. In the tables an "X" indicates that the PSU or random group was included in the half sample. The construction using 16 balanced half-sample replications results from viewing the certainty and noncertainty strata as independent universes. This is only approximately true as the post-stratified ratio adjustment to independent population controls is made across both certainty and noncertainty strata. An alternative construction, and perhaps a slightly more accurate one, would have been to use 24 bal-

Table I. Composition of the 16 balanced half-sample replicates—certainty strata

Pair	Random group of segments in certainty PSU's	Balanced half-sample replications															
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	1----------------------------------					X	X	X	X					X	X	X	X
	2----------------------------------	X	X	X	X					X	X	X	X				
2	3----------------------------------		X		X		X		X		X		X		X		X
	4----------------------------------	X		X		X		X		X		X		X		X	
3	5----------------------------------			X	X			X	X			X	X			X	X
	6----------------------------------	X	X			X	X			X	X			X	X		
4	7----------------------------------	X			X	X			X	X			X	X			X
	8----------------------------------		X	X			X	X			X	X			X	X	
5	9----------------------------------									X	X	X	X	X	X	X	X
	10---------------------------------	X	X	X	X	X	X	X	X								
6	11---------------------------------	X		X		X		X			X		X		X		X
	12---------------------------------		X		X		X		X	X		X		X		X	
7	13---------------------------------	X	X			X	X					X	X			X	X
	14---------------------------------			X	X			X	X	X	X			X	X		
8	15---------------------------------		X	X			X	X		X			X	X			X
	16---------------------------------	X			X	X			X		X	X			X	X	

Table II. Composition of the 16 balanced half-sample replicates—noncertainty strata

Pair	Sample PSU	Balanced half-sample replications															
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	Pittsburgh, Pa., SMSA---	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
	Providence, R.I., SMSA[1]																
2	Columbus, Ohio, SMSA----		X		X		X		X		X		X		X		X
	Akron, Ohio, SMSA-------	X		X		X		X		X		X		X		X	
3	York, Pa., SMSA---------			X	X			X	X			X	X			X	X
	Muskegon-Ottawa, Mich---	X	X			X	X			X	X			X	X		
4	Cayuga-Wayne, N.Y-------	X			X	X			X	X			X	X			X
	York, Me----------------		X	X			X	X			X	X			X	X	
5	Baltimore, Md., SMSA----					X	X	X	X					X	X	X	X
	Louisville, Ky., SMSA---	X	X	X	X					X	X	X	X				
6	Nashville, Tenn., SMSA--	X		X			X		X	X		X			X		X
	San Antonio, Tex., SMSA-		X		X	X		X			X		X	X		X	
7	Savannah, Ga., SMSA-----	X	X					X	X	X	X					X	X
	Midland, Tex., SMSA-----			X	X	X	X					X	X	X	X		
8	Barbour, Ala------------		X	X		X			X		X	X		X			X
	Independent cities in Virginia in 1950-------	X			X		X	X		X			X		X	X	
9	Brooks-Echols-Lowndes, Ga------------									X	X	X	X	X	X	X	X
	Jackson-Lawrence, Ark---	X	X	X	X	X	X	X	X								
10	Horry, S.C--------------	X		X		X		X			X		X		X		X
	Franklin-Nash, N.C------		X		X		X		X	X		X		X		X	
11	Lafayette-Panola, Miss--	X	X			X	X					X	X			X	X
	E. Feliciana-St. Helena, La------------			X	X			X	X	X	X			X	X		
12	San Jose, Calif., SMSA---		X	X			X	X		X			X	X			X
	Minneapolis-St. Paul, SMSA------------------	X			X	X			X		X	X			X	X	
13	Ft. Wayne, Ind., SMSA---	X	X	X	X									X	X	X	X
	Topeka, Kans., SMSA-----					X	X	X	X	X	X	X	X				
14	Grant, Wash-------------		X		X	X		X		X		X			X		X
	Apache-Navajo, Ariz-----	X		X			X		X		X		X	X		X	
15	Dunklin-Pemiscot, Mo----			X	X	X	X			X	X					X	X
	Franklin-Jackson-Williamson, Ill--------	X	X					X	X			X	X	X	X		
16	Bates, Mo---------------	X			X		X	X			X	X		X			X
	Bayfield, Wis-----------		X	X		X			X	X			X		X	X	

[1]Providence, R.I., SMSA figures into the variance computations as it is always a part of the right hand size of the difference $z'_i - z'$ in the variance equation.

Table III. Estimates of the percent of demographic subgroups of the U.S. adult population with hypertension and estimates of variance in percent

Demographic subgroup	HES estimate of percent (col. 1)	Replication estimate of variance (col. 2)	SRS estimate of percent (col. 3)	SRS estimate of variance (col. 4)	Sample persons examined (col. 5)	Ratio of variances (col. 6)
Males aged 35-44 with income less than $2,000	19.69	34.1056	22.22	27.4348	63	1.2432
Females aged 55-64 with income of $4,000-$6,999	25.75	17.2225	24.49	18.8697	98	0.9127
Females with income of $10,000+	11.75	4.7524	11.95	2.7326	385	1.7391
White males with income of $4,000-$6,999	12.21	1.2321	12.39	1.2114	896	1.0171
White females with income of $7,000-$9,999	11.48	5.7600	10.59	2.0066	934	2.8705
Negroes with income of $2,000-$3,999	23.08	10.2400	23.41	8.7474	205	1.1706
Males aged 18-24 with 9-12 years of school	2.45	1.0609	2.37	0.9151	253	1.1593
Females aged 25-34 with none or less than 5 years of school	2.49	5.3361	3.33	10.7407	30	0.4968
Males with 5-8 years of school	17.82	1.8225	17.92	1.7805	826	1.0236
White males with 13+ years of school	9.34	2.3716	9.01	1.4211	577	1.6688
White females with 9-12 years of school	10.33	0.5329	9.54	0.5187	1,666	1.0274
Negroes with 5-8 years of school	30.73	10.3684	31.19	7.4948	286	1.3834
Males aged 65-74 who are married	27.26	19.2721	26.87	9.7751	201	1.9716
Females aged 35-44 who are separated	18.80	78.3225	22.22	96.0217	18	0.8157
Females who are divorced	13.47	13.7641	14.50	9.4658	131	1.4541
White males who are single	9.05	1.9321	9.98	2.2394	401	0.8628
White females who are widowed	35.77	15.1321	33.55	7.1223	313	2.1246
Negroes who are married	27.79	5.1076	28.79	3.8316	535	1.3330
Males aged 55-64 who are craftsmen	15.15	29.9209	14.29	19.4363	63	1.5394
Females aged 35-44 who are private household workers	10.60	21.2521	10.67	12.7052	75	1.6727
Males who are laborers	19.93	11.4921	18.25	5.6730	263	2.0258
White males who are farmers or farm managers	10.89	2.8900	11.39	6.3889	158	0.4523
White females who are clerical and sales workers	9.80	2.8561	9.78	1.9522	451	1.4630
Negroes who are professional workers	16.57	31.2481	16.22	36.7202	37	0.8510
Males aged 25-34 who are employed in construction and mining	7.46	9.3025	9.52	13.6775	63	0.6801
Females aged 18-24 who are employed in wholesale and retail trade	2.29	6.8121	2.27	5.0477	44	1.3495
Males who are employed in transportation	11.29	4.1209	10.76	4.3067	223	0.9569
White males who are employed in finance, insurance and real estate	12.34	24.5025	11.54	13.0860	78	1.8724
White females who are employed in services	10.48	3.0276	10.59	2.3324	406	1.2981
Negroes who are employed in Government	22.19	9.3636	22.65	9.6800	243	0.9673

Table IV. Estimates of differences in percent between demographic subgroups of the U.S. adult population with hypertension and estimates of variance in percent

Demographic subgroup	HES estimate of percent (col. 1)	Replication estimate of variance (col. 2)	SRS estimate of percent (col. 3)	SRS estimate of variance (col. 4)	Sample persons examined (col. 5)	Ratio of variance (col. 6)	Average of replicate percents (col. 7)
1. Adults with income less than $2,000-------	26.23	4.5796	26.66	1.7791	1,099	2.5741	26.06
2. Adults with income of $10,000+------------	11.75	1.3924	12.17	1.3933	764	0.9994	11.34
Difference (1-2)-----	14.48	5.2850	14.48	3.1784		1.6628	15.72
3. Males with income $2,000-$3,999----------	15.36	4.2025	15.51	2.4499	535	1.7154	14.75
4. Males with income $4,000-$6,999----------	12.83	1.4641	13.24	1.1797	975	1.2411	12.35
Difference (3-4)-----	2.53	4.5600	2.27	3.6296		1.2563	2.40
5. Females aged 55-64 with income $4,000-$6,999---	25.75	17.2225	24.49	18.8697	98	0.9127	25.91
6. Females aged 55-64 with income $7,000-$9,999---	25.25	99.8001	24.32	49.7500	37	2.0060	26.67
Difference (5-6)-----	0.50	110.1069	0.17	68.6197		1.6046	0.66

anced half-sample replications, viewing the 16 pairs of noncertainty strata and 8 pairs of randomly grouped clusters from the certainty strata as a single universe.

After the composition of each of the balanced half samples was determined, the resulting half samples were then separately subjected to all the estimation procedures and tabulations used to produce the final estimates from the entire sample.

An estimated variance $s^2_{z'}$ of an estimated statistic z' of the parameter z is obtained by applying the formula

$$s^2_{z'} = \frac{1}{16} \sum_{i=1}^{16} (z'_i - z')^2$$

where z'_i is the estimate of z based on ith half sample and z' is the estimate of z based on the entire sample.

Computer Output

For the Health Examination Survey the variance tabulations and prepublication tabulations of estimates are derived from the same computer output. Since the findings are generally expressed as rates, means, or percentages, each output "table" actually consists of three tables, the statistic of interest, such as the percent of persons with hypertension, the numerator of each cell in the "table," and the denominator of each cell. The cells of the table are a cross-classification of the statistic by age and sex with one of about a dozen demographic variables for which information was collected in the survey. The analyst can also receive a printout of the same three tables for each of the 16 half-sample replications. The replication tables are useful when estimates of the variance of estimated differences between statistics or of such derived statistics as medians are needed or for evidence to support or refute a hypothesis concerning observed patterns in the data. In addition to the "table" of findings, the output also includes a "table" of estimated standard errors (of the statistic, its numerator, and its denominator), a "table" of estimated relative variances (the estimated variance of a statistic divided by the square of the statistic), and a "table" of the number of sample observations on which the statistic, its numerator, and its denominator are based. The last table together with the others gives some insight into the effect of the sampling plan and estimation procedures.

Illustration

The figures in table III are estimates from the Health Examination Survey of the percent of demographic subgroups of the adult population with hypertension and their estimated variances. The official HES estimates based on unbiased inflation factors adjusted for nonresponse and ratio adjusted to independent population controls are shown in column 1. Estimates of their variance derived from 16 balanced half-sample replications treating the estimated percent of replicate i as z_i' are shown in column 2. For comparison, the estimates of percent and variance which would have resulted if the 6,600 examined persons had been a simple random sample of the U.S. population and the sample size in each demographic subgroup or domain is considered to be fixed, are shown in columns 3 and 4. The number of examined sample persons in the demographic subgroup or domains (the bases of the percents) are shown in column 5. The ratios of the two variance estimates are shown in column 6. These ratios are indicative of the net effect of clustering and stratification in the sample design, deviations from equal probabilities of selection, and nonresponse and ratio adjustment in the estimation procedures, and reflect as well the variance of the estimated variance.

The median ratio of replication variance to simple random variance—i.e., of an appropriate variance to a much cruder measure—is 1.30. The mean ratio is 1.31. As one would expect, there is a tendency for the ratio to be higher for larger values of the statistic, although this tendency is not very pronounced.

The criteria for hypertension was 160 mm. Hg. or over systolic blood pressure and 95 mm. Hg. or over diastolic. The average of three blood pressures taken over a 30-minute period was used for each examined person.

Table IV is similar to table III but it also includes the estimated difference in percent between two demographic subgroups. Estimates of variance of the difference between two estimated percents which would have resulted if the sample had been a simple random sample were obtained by summing the estimated variances of the two estimated percents. The average of the estimated percents over the 16 replicates is shown in column 7.

———o o o———

*U.S.GOVERNMENT PRINTING OFFICE : 1979 0-281-259/30

VITAL AND HEALTH STATISTICS Series

Series 1. *Programs and Collection Procedures.*—Reports which describe the general programs of the National Center for Health Statistics and its offices and divisions and data collection methods used and 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.

Series 3. *Analytical 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 10. *Data From the 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 a continuing national household interview survey.

Series 11. *Data From the Health Examination Survey and the Health and Nutrition Examination Survey.*—Data from direct examination, testing, and measurement of national samples of the civilian noninstitutionalized population provide the basis for two types of reports: (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 effective 1975. Future reports from these surveys will be 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 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 effective 1975. Future reports from these sample surveys based on vital records will be 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 biennial 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

DHEW Publication No. (PHS) 79-1269
Series 2-No.14

NCHS

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
Public Health Service
Office of Health Research, Statistics, and Technology
National Center for Health Statistics
3700 East West Highway
Hyattsville, Maryland 20782

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE, $300

POSTAGE AND FEES PAID
U.S. DEPARTMENT OF H.E.W.
HEW 396
THIRD CLASS

For publications in the *Vital and Health Statistics* Series call 301-443-NCHS.