THE SURVEY OF INCOME AND PROGRAM PARTICIPATION

VARIANCE ESTIMATION BY USERS OF SIPP MICRO-DATA FILES

No. 9303 175

Rameswar P. Chakrabarty

Statistical Research Division
Bureau of the Census

U.S. Department of Commerce BUREAU OF THE CENSUS

VARIANCE COMPUTATION BY USERS OF SIPP MICRO-DATA FILES

INTRODUCTION

Public use data files of the Survey of Income and Program Participation (SIPP) are expected to be used by many different researchers, particularly in the areas of Sociology, Economics and public policy. Possible uses of SIPP Micro-Data Files can be broadly classified into two major areas: (1) computation of summary statistics such as means, totals and ratios for states, regions, and subpopulations like Blacks, Hispanics, low income families, etc. (beyond what is published by the Bureau of the Census in SIPP reports), and computation of variances of such statistics for tests of hypotheses and statistical inferences, (2) analytical studies to understand various socio-economic phenomena such as factors affecting the dependency on welfare, or variables affecting the risk of experiencing an event; e.g., marriage, child birth, divorce or unemployment. These analytical studies will generally involve some form of multivariate analysis and statistical modeling techniques. Data analysts can usually compute summary statistics from micro-data files easily. Estimation of variances and multivariate analysis of survey data are, however, not so easy. It is, therefore, desirable to provide guidelines to users on how to compute sampling errors and conduct a multivariate analysis or statistical modeling. The guidelines should also indicate software packages that are appropriate for computation of variances or multivariate analyses so that users can carry out these complex statistical computations relatively easily.

Exploratory research was undertaken to review general methods of variance estimation from complex surveys and multivariate analysis of survey data, and all available software packages with regard to their suitability for analysis of data from the complex sample design of the SIPP. This research mainly deals with cross - sectional analysis of data. No general purpose software packages are available for longitudinal analysis of data from complex surveys. This report provides guidelines for variance estimation. Multivariate analysis of SIPP data has been discussed in another report by Chakrabarty (1989).

General methods of variance estimation from complex surveys with their advantages and disadvantages are discussed first. This is followed by a discussion of all available software packages with regard to their statistical methodology, capabilities, computational complexities, and most importantly, their suitability for computing variances from the complex sample design of the SIPP, along with some recommendations.

1. GENERAL METHODS OF VARIANCE ESTIMATION FROM COMPLEX SURVEYS

Three general methods are available for computing the sample estimates of the variances of non-linear statistics like ratio estimate, and simple, partial and multiple regression and correlation coefficients. These are Taylor Series Linearization (TSL), Balanced Repeated Replications (BRR) and the Jackknife Repeated Replications (JRR). The essential features of these methods are outlined below. A more detailed discussion of these methods can be found in Wolter's (1985) book. Some recent developments in the properties of these methods for complex surveys are given in Rao (1988), Rao and Wu (1988), Rao, Wu and Yue (1992) and Rao (1993).

1.1 Taylor Series Linearization (TSL)

Let $t=(t_1,t_2...t_k)$ be a set of statistics whose expected value is the set $T=(T_1,T_2...T_k)$. If the function to be estimated $\theta=F(T)$ is estimated by $\hat{\theta}=F(t)$, then

$$\hat{\theta}-\theta=F(t)-F(T)=\sum_{i=1}^{k}(t_i-T_i)\frac{\partial F}{\partial T_i}$$

approximately, using the first term Taylor expansion. The partial derivatives are to be evaluated at $t_i = T_i$ but in practice usually have to be evaluated at t_i. The variance of $\hat{\theta}$ is then approximated by the variance of the linear function

$$\sum_{i=1}^{k}(t_i-T_i)\frac{\partial F}{\partial T_i}$$

given by

$$V(\hat{\theta})=\sum_{i=1}^{k}\left(\frac{\partial F}{\partial T_i}\right)^2 V(t_i)+\sum_{i\neq j}\frac{\partial F}{\partial T_i}\,\frac{\partial F}{\partial T_j}Cov(t_i,t_j).$$

This is the method which produces the usual large-sample formula for the variance of the ratio estimate given in literature (see Kendall and Stuart 1963 and Cochran 1977). In extending this method to complex surveys and estimates, papers by Keyfitz (1957), Kish (1968), Tepping (1968), Woodruff (1971), and Woodruff and Causey (1976) have shown that ingenuity in the method used to compute the variance of the linearized form can considerably simplify the computation. Some underestimation of variance is to be expected at least for moderate-sized samples because higher order terms are neglected in TSL. Sukhatme, et al. (1984) expressed the error in the variance in terms of the moments of a bivariate distribution and provided some upper bounds for the errors in the approximation. The underestimation of the variance of the ratio estimate by this method has been confirmed recently (See Rao 1968, 1969, Mulry and Wolter 1981, Krewski and Rao 1981, and Efron 1982). Dippo and Wolter (1984) apply second order Taylor Series approximations to several estimates and show that this reduces the bias, but increases the variance of the variance estimate and complicates computations considerably. Underestimation of the variance from Taylor series may compensate for some over estimation that results when variance is computed assuming sampling with replacement for a without replacement sampling design.

1.2 The Jackknife Repeated Replications (JRR)

The Jackknife method was first suggested by Quenouille (1956) for reducing the bias of an estimator like the ratio (based on a simple random sample of size n) from order 1/n to order $1/n^2$. Suppose n=mg; the sample is divided into g groups of size m. The jackknife estimator of θ is

$$\hat{\theta}_J = \frac{1}{g}\sum \hat{\theta}_j$$

where $\hat{\theta}_j = g\hat{\theta}_s - (g-1)\hat{\theta}_{(s-j)}$; $\hat{\theta}(s-j)$ denote the estimate of θ made by omitting the jth group and $\hat{\theta}_s$ is the estimate from the complete sample.

Tukey (1958) suggested that in many nonlinear situations, the variance of the jackknife estimator $\hat{\theta}_J$ can be estimated by

$$\hat{V}(\hat{\theta}_J) = \sum_{j=1}^{g} (\hat{\theta}_j - \hat{\theta}_J)^2 / g(g-1)$$

treating pseudo values $\hat{\theta}_j$ as though they are independent estimates. He also suggested that

$$t' = (\hat{\theta}_j - \theta) / \{\hat{V}(\hat{\theta}_J)\}^{1/2}$$

would often be approximately distributed as student's t with (g-1) d.f. With linear estimates from normal data t' is identical to t.

Some analytic support for this approach comes in asymptotic as well as *exact* results. Under broad assumptions, Brillinger (1964) showed that for maximum likelihood estimates, t' tends to t distribution with (g-1) d.f. as m tends to infinity. Arvensen (1969) gave a similar result for estimates (of the form known as Hoeffding's U Statistics) that are symmetrical with respect to members of the sample. In a Monte Carlo study, Rao and Beagle (1966) found that the stability of the jackknife variance estimator of a ratio estimate is similar to the TSL variance estimator. For analytical results that are *exact* for any sample size n, Chakrabarty and Rao (1967) assume a linear regression model where auxiliary variable x has a gamma distribution and show that the *exact* bias of the jackknife variance estimator decreases as g increases and with g=n the bias is less than the TSL variance estimator for moderate sample sizes. The *exact* stability of the variance estimator with g=2 (i.e., 2 replications) is shown to be less than the TSL variance estimator. In an evaluation study using seventeen natural populations, Rao and Kuzik (1974) found that the stability of the jackknife variance estimator (with g=n) to be comparable to that of the TSL variance estimator. Kreswki and Chakrabarty (1981) later show that the *exact* stability of a slightly modified jackknife variance estimator with g=n is comparable to that of the TSL variance estimator under conditions specially favorable to ratio estimation (i.e., regression approximately through the origin with a relatively small coefficient of variation in the x population).

1.3 Balanced Repeated Replications (BRR)

The Balanced Repeated Replications (BRR) has mostly been developed for stratified samples with 2 PSUs per stratum. With 2 PSUs per stratum chosen with equal probability, a half-sample is obtained by selecting at random one PSU from each stratum.

Let $\hat{\theta}_H$, $\hat{\theta}_c$ and $\hat{\theta}_s$ denote the estimates computed from the chosen half-sample, the complimentary half-sample, and the whole sample. For strictly linear estimate, the quantities

$$(\hat{\theta}_H-\hat{\theta}_s)^2=(\hat{\theta}_c-\hat{\theta}_s)^2=\frac{1}{4}(\hat{\theta}_H-\hat{\theta}_c)^2 \qquad (1)$$

are all unbiased estimates of $V(\hat{\theta})$, if the finite population correction is negligible. With only two half-samples, the reliability of the variance estimate is poor but can be increased by increasing the number of replications. With L strata, the possible number of replications is 2^L. For linear estimates, McCarthy (1966) has shown that an orthogonal subset, using the Plackett & Burman (1946) orthogonal main effects for the 2^L factorial, produces the same variance estimate as the complete set, and requires at most (L+4) half-replicates. McCarthy's balanced half-sample replications (BHR) method is often called balanced repeated replications (BRR). If the number of balanced half-replicates is too large, McCarthy suggests use of a subset of the balanced set. In a later review of this method, McCarthy (1969) found that the half-sample variance estimators in (1) agreed well with one another also for some non-linear estimates -- 15 combined ratio estimates, 24 partial regression coefficients and 8 multiple regression coefficients for the Health Examination Survey. The National Center for Health Statistics (NCHS) uses this method for variance computation in the Health Interview Survey (HIS) and other health surveys. Prior to the development of Balanced Half Samples, the repeated replications were achieved by dividing the sample at random into several subsamples or replications. This method called Random Groups (RG) was used extensively for variance estimation in the CPS in the 1950's and early 60's and in various business and market research surveys by Deming (1956,1960). It is still being used in Business Surveys by the Bureau of the Census. Historically, this was one of the first techniques developed to simplify variance estimation. It was introduced by Mahalanobis (1939, 1944) who used "interpenetrating subsamples" as replications. The Random Groups method and its properties are given in

Wolter (1985). In designs with few PSUs per stratum it becomes necessary to collapse strata to form enough random groups to obtain a reliable estimate of variance. Collapsing strata, however, defeats the objective of obtaining an unbiased estimate of variance for a 2 PSUs per stratum design.

1.4 Recent Developments: Bootstrap and Replicate Weights

Recently, a new method of variance estimation called "Bootstrap" has been suggested by Efron (1982). This method is as follows: From n observations $y_1\ y_2 \ldots y_n$ drawn independently from an unknown probability distribution F, fit the nonparametric maximum likelihood estimate of F. It can be simply the uniform weight 1/n at each value y_i (i=1,2...n). Draw m bootstrap samples of size n_1 (< n) from this estimated distribution independently. The variance of the estimate $\hat{\theta}$, based on the full sample, is estimated by

$$\hat{V}(\hat{\theta}) = \sum_{k=1}^{m} (\hat{\theta}_k - \hat{\theta})^2/(m-1)$$

where $\hat{\theta}_k$ is the estimate from the *k*th bootstrapped sample. Efron shows that the bootstrap estimator slightly underestimates the variance. This method like other resampling methods will be computer intensive. A simulation study by Kovar, Rao and Wu (1988) indicates that this method performs well in estimating confidence intervals for parameters but the variance estimator is less stable than those based on TSL or JRR. Rao, Wu and Yue (1992) provided an extension of the bootstrap method to stratified multistage designs, and compared the bootstrap variance estimator for the median with the BRR method by simulation. The bootstrap variance estimator had a larger relative bias and a larger coefficient of variation compared to the BRR variance estimator. At this time we do not know of any successful application of the bootstrap method to complex survey data.

Fay (1984) has developed a replicate weighting representation of the replication method. The general method of replicate weighting is to assign to each sample case i replicate weight W_{ir}, r=1,...R (R is the number of replicates) so the estimate of the population total Y from the *r*th replicate is

$$\hat{Y}_r = \sum_i W_{ir} y_i$$

The estimate of Y from the whole sample is

$$\hat{Y}_o = \sum_i W_i y_i$$

where W_i is the final weight for *i*th sample case. The variance estimator used with these replicate estimates is typically of the form:

$$\hat{V}(\hat{Y}_o) = \sum_r d_r (\hat{Y}_r - \hat{Y}_o)^2$$

where replicate factors d_r, r=1,...R are independent of the choice of characteristic y but may possibly depend on the selected sample and r. For many common replication methods d_r will simply be a constant depending on R, such as 1/R or (R-1)/R.

The basic replicate weight for the sample case i in the *r*th replication is based on the survey design and the inverse of probability of selection. These weights may be constructed according to a familiar replication method such as BRR or according to a more general resampling plan. Final replicate weights W_{ir} are formed according to the estimation procedure (nonresponse adjustments, first and second stage ratio adjustments, etc). Thus, it reweights each replicate to population control totals. This procedure generally provides a conservative estimate of variance in the sense that it overestimates variance. The procedure, however, is very flexible and can be geared to specific survey designs by obtaining the correct replicate factors. Most importantly, this method like other resampling methods, permits the computation of design-based estimates of variance for a wide variety of statistics including analytically complex statistics for which TSL may be difficult.

Dippo, Fay and Morganstein (1984) report their favorable experiences of using some variations of this method in several surveys including the 1984 SIPP. Fay (1989) provides the theory and application of his method in the 1985 panel of SIPP.

1.5 Comparison of The Three Methods

The *exact* analytic comparison of the three methods appears to be formidable; not much is known beyond what was mentioned earlier about the model-based comparison of the *exact* bias and stability of JRR and TSL variance estimators by Chakrabarty and Rao (1967), and Kreswki and Chakrabarty (1981). Research on comparison of these methods has been largely empirical: Frankel (1971), Kish and Frankel (1974), Bean (1975), Campbell and Meyer (1978), Lemeshow and Levy (1978), Shah et al (1977) and others. Recently, asymptotic analytic comparisons have been made by Krewski and Rao (1981), Dippo and Wolter (1984) and Rao and Wu (1985). A description of some of these studies and important results follow.

Frankel (1971), and Kish and Frankel (1974), made a Monte Carlo comparison of the performance of three methods in small samples using the March 1967 CPS data as a population. The population contained 3240 primary units of average size 14.1 households and were divided into 6, 12, and 30 strata. Two units were drawn from each of 6, 12 and 30 stratum providing samples of sizes 12, 24 and 60. Thus, the study involved cluster units of unequal sizes and proportional stratification, but not unequal probability selection.The types of estimate examined were: 8 means (that were ratio estimates), 12 differences of means, 12 simple correlations, 8 partial regression coefficients, 8 partial correlations and 2 multiple correlations. As regards the properties of the estimators, the average relative biases of the estimators, $[E(\hat{\theta})-\theta]/\theta$, were less than 1% for means and differences between means, less than 7% for simple correlation and 5% for partial regression coefficients even with the smallest sample size 12. The average relative biases were somewhat larger for partial and multiple correlations (12% and 16% for n=12). Examination of the ratio of these biases to the standard error of the estimate showed that in the worst cases, two sided 95% confidence interval statements might have confidence probability nearer 90%.

Frankel compared four variations of both the BRR and JRR with the TSL variance estimator. No method is consistently best in all situations. All three methods do satisfactorily well for means and for differences between means, the average biases being generally under 5%, and the bias of JRR is generally least. BRR is the only method that does well for simple correlations and is superior to JRR for partial correlations (TSL was not tried for partial and multiple correlations because of its complexity in these cases). On the other hand, both TSL and JRR are superior to BRR for partial regression coefficients.

Turning to inferences about parameters, Kish and Frankel (1974) evaluated the empirical coverage probability of the $1-\alpha$ level confidence intervals, $\hat{\theta}\pm t_{\alpha/2}\{\hat{V}(\hat{\theta})\}^{1/2}$, where $t_{\alpha/2}$ is the upper $\alpha/2$-*point* of the student's t distribution, for means (ratios) differences between means, regression and correlation coefficients. The three methods seemed to be satisfactory for means, differences between means and regression coefficients. The BRR method performed better than the JRR method which in turn did better than the TSL method in terms of coverage probability. The three methods, however, performed in the reverse order in terms of the stability of variance estimator; the mean square error (MSE) of the TSL variance estimator was smallest. The best BRR method for variance estimation,

$$\hat{V}(\hat{\theta}_s) = \text{ average value of } [1/2(\hat{\theta}_H-\hat{\theta}_s)^2+1/2(\hat{\theta}_c-\hat{\theta}_s)^2](1-f),$$

where (1-f) is the finite population correction, was most satisfactory for inferences about means, differences between means and regression coefficients. But

$$\hat{V}(\hat{\theta}_s) = \text{ Average value of}$$

which does not require calculation of $\hat{\theta}_c$, does almost as well. The inferences about simple, partial and multiple correlations were not very satisfactory and sometimes erratic as agreement between t' and t worsens as n increases. This might have been due to skewed distribution of correlations bounded between -1 and +1 or due to small sample sizes of 12, 24 and 60 in this Monte Carlo study.

Bean (1975) compared BRR and TSL methods for computing variances of ratio estimates in a Monte Carlo study using data from the Health Interview Survey. Both methods gave satisfactory variance estimates and adequate two sided confidence probabilities but one sided confidence intervals were not reliable. Further empirical investigations of properties of these methods by Campbell and Meyer (1978), Lemeshow and Levy (1978), Shah, Holt and Folsom (1977), Rust (1985), and Anderson et al. (1987) provided similar results.

For nonlinear statistics that can be expressed as functions of estimated totals, Krewski and Rao (1981) established asymptotic consistency of TSL, JRR and BRR variance estimators as number of strata becomes large. Their first-order asymptotic result is valid for any stratified multi-stage design in which PSUs are selected with replacement and in which independent subsamples are selected within those PSUs sampled more than once. Analytical properties of TSL, JRR and BRR variance estimators are, however, indistinguishable in their expansions to first order terms only. Dippo and Wolter (1984) compare RG, JRR and BRR analytically by including all second order terms in their Taylor series expansions. The Consumer Expenditure Survey (CES) diary data (1980-81) is used to evaluate variance estimators in small samples for extremely skewed distributions of many CES variables. The results of this study indicate that the bias of RG, BRR and JRR variance estimators is small; the relative bias is less than 9%. The variance of the variance estimators, on the other hand, is not insignificant and the normal theory confidence intervals do not always have the desired coverage probabilities. Users should note that the construction of confidence intervals and tests of hypotheses assuming normality may not be appropriate in situations where the numerator of the ratio estimator is a function of a variable from a very skewed population. Confidence intervals based on transformations may be better in these situations as shown in Mulry and Wolter (1981). Rao and Wu (1985) have made an asymptotic second order comparison of the BRR, JRR and TSL methods for any stratified multi-stage design in which PSUs are selected with replacement. When the design consists of two sampled PSUs per stratum, the TSL variance estimator is shown to be identical (in second-order asymptotic expansions) to the BRR variance estimator (that uses both $\hat{\theta}_H$ and $\hat{\theta}_c$) and to the JRR variance estimator (called $V_{JRR\text{-}D}$ in Kish and Frankel 1974) for nonlinear estimates such as ratio, correlation and regression coefficients. These results suggest that for 2 PSUs per stratum designs with large number of strata, there is not much to choose between TSL, BRR and JRR variance estimators in terms of statistical criteria and therefore, practical considerations such as available computing resources and computing costs should dominate the choice of a variance estimator.

In practical applications of variance estimation methods, it should be noted that: (1) replication or resampling methods - BRR, JRR and Bootstrap require more extensive computation than the TSL method and will continue to be "computer-intensive" (Diaconis and Efron 1983); (2) some sample designs may not satisfy the restrictions required by the replication methods. The TSL method is, however, applicable to any sample design; (3) replication methods may facilitate the estimation of variance for very complex functions for which TSL would require derivation of new variance formulas.

2. SOFTWARE PACKAGES.

Computer software packages for the analysis of complex sample survey data and variance estimation are being developed by many institutions. At present, there are about 14 software packages available for computing variances from complex surveys (See APPENDIX). Some of these are not easily accessible, others are limited in scope or developed for a specific sample design. Our review indicates that of these fourteen, only three are versatile, well supported, portable programs, and applicable to stratified multi-stage sample designs. These are SUDAAN, SUPER CARP, and OSIRIS IV developed by the Research Triangle Institute (RTI), Iowa State University and the Survey Research Center, University of Michigan respectively. Note that the enhancements to these three programs as well as to several other programs are in progress. New developments could make other programs, particularly TREES and NASSTIM, worth considering. The key features of SUDDAN, SUPER CARP and ORSIS IV are outlined below.

2.1 SUDDAN: Survey Data Analysis

Developed and distributed by
Research Triangle Institute
Post Office Box 12194
Research Triangle Park
North Carolina 27709

Source Languages.

Fortran and Assembler

Compatibility

The program will run on any IBM system on which Statistical Analysis System (SAS) has been implemented.

Documentation

Holt, M.M. (1977). SURREGR: Standard Errors of Regression Coefficients from sample survey data. Research Triangle Institute, N.C.

Lavange, L.M., Shah, B.V., Barnwell, B.G., and Killinger, J.F. (1989). SUDAAN: A Comprehensive Package for Survey Data Analysis, Technical Report. Research Triangle Institute N.C.

Shah, B.V. (1976). STDERR: Standard Errors Program for sample survey data. Research Triangle Institute, N.C.

Shah, B.V. (1981). RATIOEST: Standard Errors Program for Computing of the Ratio Estimates from sample survey data. Research Triangle Institute, N.C.

Shah, B.V. (1981). SESUDAAN: Standard Errors Program for Computing of Standardized Rates from sample survey data. Research Triangle Institute N.C.

Shah, B.V. (1982). RTIFREQS: Program to Compute Weighted Frequencies, Percentages and their standard errors. Research Triangle Institute, N.C.

Description

B.V. Shah has developed software at the Research Triangle Institute (RTI) for analyzing data collected from surveys with stratified multi-stage sample designs. These procedures can be accessed only through the Statistical Analysis System (SAS). The RTI intends to provide continued support for the software and has long range plans to develop a survey data analysis (SUDAAN) language that will greatly facilitate analysis of survey data.

The SUDAAN procedures compute means, proportions, ratios and regression coefficients and their variances. The statistical approach used for computing the standard error is the Taylor series linearization (TSL) method.

The SUDAAN at present has four procedures. SESUDAAN, RATIOEST, RTIFREQS and SURREGR. A synopsis of these procedures follows:

SESUDAAN: Standard Errors of Means, Proportions, Standardized Rates and Differences

The program provides estimates of means or proportions for many domains (subgroups of populations) of interest. The domains are conveniently specified by the user with the TABLES statement similar to the one in PROC FREQ, SAS.

The major conveniences for comparative evaluation are standardization and differencing. For example, for comparing mortality rates of the states of Florida and New York, it would be appropriate to compute rates standardized to population for both states and then take the difference between the two. SESUDAAN permits computation of such standardized differences and their standard errors for many domains.

Other convenience features include: (a) computation of design effects and various sample sizes, (b) output of results as SAS data sets, and (c) flexible format of tables. SESUDAAN is very efficient and runs approximately three times faster for equivalent tasks than on STDERR, an older program.

RATIOEST: Program to Compute Standard Errors of Ratios

This program is similar to SESUDAAN except for the following differences:

(a) Standard errors are computed from statistics of the form $\Sigma WY/\Sigma WX$ instead of $\Sigma WY/\Sigma W$

(b) Standardization or differencing is not permitted.

(c) Standard errors of the totals are not estimated.

RTIFREQS: Program to Compute Standard Errors of Percentages and Estimated Population Totals

This program is very similar to SAS procedure FREQ. It computes sample counts, estimated totals, and percentages (row, column and total) as well as their standard errors.

SURREGR: Standard Errors of Regression Coefficients from Sample Survey Data

SURREGR is a procedure which provides a means of producing appropriate tests of hypotheses for regression models in sample survey situations. The procedure offers many useful options and operates in three modes which differ only in the method by which the variance-covariance matrix of the regression coefficients is calculated. SURREGR was primarily developed to handle regression analysis for sample survey data; hence, the default mode of the procedure will incorporate a stratified multistage sampling design into the variance-covariance computation. Another mode of the procedure relies on the ordinary least squares estimate for the variance-covariance matrix. Also, a weight may be used for a weighted ordinary least squares analysis. Some useful options for SURREGR are described below.

- The DATAOUT option produces a SAS file which contains for each model, the regression coefficients, the variance-covariance matrix, the F test values, and their associated degrees of freedom.

- The RESIDUAL option allows for output to a SAS data set of the unweighted predicted and residual values associated with each level of each dependent effect for each model.

- The BETA option prints a solution to the normal equations and the variance-covariance matrix for the solution.

- The MODEL statement allows the user to list one or more multiple dependent effects with any number of independent effects. An effect may be a single variable or a main effect, or it may be composed of a group of variables. When there is more than one variable in an effect, each variable must be joined to the next with either a * indicating crossed variables or a () indicating a nesting structure. An effect may contain continuous or discrete variables, but only discrete variables may be nested. Variables which are combined into one effect must be listed with the crossed and then the nested groupings. Only one level of nesting is allowed. For computational aspects and empirical studies related to the approximation, the reader is referred to Shah, et al. (1977); Holt (1977); and Fuller (1975).

- Finally, note that significant improvements to the SUDAAN package are provided in the new SUDAAN system under development (Lavange et al. 1989).

2.2 SUPER CARP: Cluster Analysis and Regression Program

Developed and Distributed by
Survey Section
Statistical Laboratory
Dept. of Statistics
211 Snedecor
Iowa State University
Ames, Iowa 50010

Compatible with following computer systems:
IBM 360 and 370
Univac 1100

Operating Systems
OS, TSO
Source Languages
Fortran G

Documentation

Hidiroglou, M.A., Fuller, W.A. and Hickman, R.D. (1980). SUPER CARP, Iowa State University, Ames, Iowa

Schnell, D., Kennedy, W.J., Sullivan, G., Park, H.J., and Fuller, W.A. (1988). Personal Computer Variance Software for Complex Surveys, Survey Methodology, 14, 59-69.

Description

SUPER CARP (Cluster Analysis and Regression Program) was written by Hidiroglou, Fuller and Hickman at the Iowa State University. This program computes estimates of population (and subpopulation) means, totals, proportions and ratios and their variances for stratified multi-stage samples. The variances of means, totals and ratios are obtained by Taylor series approximation. For PPS sampling with 2 PSUs per stratum, the program computes the variance using the Yates-Grundy formula which under estimates the total variance whenever sampled PSUs are subsampled (see Raj 1968).

SUPER CARP computes regression coefficients and their standard errors using the methods given in Fuller (1975). A special feature of this software provides regression estimates in the presence of known (or estimated) response errors. A test of goodness of fit and test of independence in a two way table are provided in addition to test for regression coefficients. SUPER CARP also allows for automatic collapsing of strata that contain only one PSU. Each one PSU stratum is combined with the next stratum in the sequence. Note that this may not be an optimal procedure. It provides standard errors for stratum means, totals and ratios.

The maximum number of variables that can be included in the standard program in one analysis is 50. If finite population correction option is elected for the two-stage sampling designs, the maximum number of PSUs is 2000.

A significant enhancement to SUPER CARP is the introduction of PC CARP (Schnell et al 1988), available on IBM AT/XT or compatible micro-computers with a math co-processor. This package, like SUPER CARP, uses the TSL method for variance estimation.

2.3 OSIRIS IV

Developed and distributed by
The Institute for Social Research
The Survey Research Center
University of Michigan
Ann Arbor, MI 48106

Compatible with following computer systems.

IBM 360/370
IBM Compatible such as
AMDAHL 470 V/6 Operating systems:
MTS
OS/360
MVS or equivalent

Source Languages:

Fortran IV or Assembler

Documentation

Computer Support Group (1982). OSIRIS VI: Statistical Analysis and Data Management Software System. Survey Research Center. Institute of Social Research, University of Michigan.

Computer Support Group (1980). Sampling Error Analysis in OSRIS IV. Survey Research Center, Institute for Social Research, University of Michigan.

Description

The OSIRIS IV software system was written by Kish, Frankel and Van Eck at the Institute for Social Research, University of Michigan. It has two programs for computing sampling error estimates derived from surveys with complex sample designs. The PSALMS command produces sampling error estimates for ratios, ratio means, totals and differences of ratios. The REPERR command produces sampling error estimates for means and regression coefficients based on replication method of variance estimation. Three alternative forms of replication methods are available: Simple replications, balanced repeated replications (BRR), and jackknife repeated replications (JRR).

For computational ease and generality, both commands assume that the PSUs are selected with replacement. Note, however, that this method leads to an overestimation of variance if PSUs are selected without replacement.

Sampling error estimation requires at least two PSUs within each stratum, When the sample design does not provide this, the collapsed strata technique (Cochran 1977) needs to be applied. Required input specifications for both commands include the definition of a stratification variable and a sampling error computing unit (SECU) variable.

The definition of the SECU variable is fundamental to the estimation procedure. With multi-stage sampling, each PSU could constitute a separate SECU; with single-stage sampling, each unit can be a SECU. If PSUs are numerous and small, it may be advantageous to join several PSUs to form a SECU. PSUs may be combined within a stratum when a stratum contains many PSUs; alternatively, they may be combined across strata using the 'thickening zone' technique (Deming 1960). The reliability of variance estimates increases as the number of SECUs increases. The amount of computing in the PSALMS command is relatively independent of the number of SECUs and, therefore, as many SECUs as possible should be formed (each PSU should be a separate SECU). The amount of computations in the REPERR command is dependent on number of replications formed, which in turn depends to a certain extent, on the number of SECUs formed, particularly with the BRR method. Therefore, it is sometimes desirable to employ a smaller number of SECUs for the REPERR command.

The PSALMS command uses the Taylor series expansion method to estimate variances for ratios, means, totals and differences of ratios. It enables a single run to yield sampling errors for a range of estimates for both the total sample and an unlimited number of subpopulation domains. The statistic, its standard error and simple random sampling standard error, design effect, intraclass correlation, and unit and weighted counts are included in the printout. A special feature of this is that the printout includes coefficients of variation of the denominator of ratios and a warning when it is determined to be too high, indicating inadequacy of the Taylor series approximation for large samples.

The REPERR command computes variances for means, correlations, regression coefficients, standardized regression coefficients and multiple correlation coefficients using replication techniques. Replications are formed using BRR, JRR or from user-specified lists of SECUs. Only one model may be used in a run. The BRR model requires that each stratum contain exactly two SECUs, one SECU is selected from each stratum to form a replication. This operation is repeated to form a set of replications that have the property of orthogonality (Kish and Frankel 1974, Plackett and Burman 1946). The number of strata that can be accommodated by this procedure is limited from 4 to 88. The compliments of replications are also included as replications and used in variance computations, as discussed by Frankel (1971). The JRR model can be applied with any number of SECUs per stratum greater than one. The standard printout includes the statistic, its standard error, simple random sample error, design effect and test statistics (corresponding to the t-statistics under assumptions of simple random sampling and normality). Printout options available include: the listing of SECUs and weights by replication; SECU and replication unvariate statistics, sums, sums of squares and products; and regression analysis by replication and for the total sample. A feature is also available to create dummy variables from categorical measures for use as independent variables in regression analysis (See Draper and Smith 1966, and Kish and Frankel 1974).

2.4 Comparison of Software Packages.

Recently, Cohen, Burt and Jones (1986) have compared four variance programs: SESUDAAN/RATIOEST, in the SUDAAN package, SUPER CARP, OSIRIS/PSALMS, and HESBRR[1] using data from the National Medical Care Expenditure Survey. The comparisons concentrate on program capabilities, computational efficiency, and user facility. We discussed program capabilities earlier. Results for user facility and computational efficiency are given in tables 1 and 2 respectively. Table 1 compares the ease of implementation of software packages in terms of number of programming statements required to implement a program for computation of variances of means, totals, and ratios. Table 2 compares computational efficiencies of software packages in terms of CPU time needed to compute variances of means, totals and ratios for three different sample sizes.

TABLE 1
Ease of Applications - Number of Programming Statements

	Number of Programming Statements			
Variance of	SESUDAAN/ RATIOEST	OSIRIS/ PSALMS	SUPER CARP	HESBRR 1/
Means	9	144	50	168
Totals	9	144	50	168
Ratios	10	354	116	420

The SESUDAAN procedure required only 9 program statements to yield variances for means or totals and the RATIOEST procedure required only 10 program statements to obtain variances for ratios. Clearly, the SESUDAAN/RATIOEST procedure was superior in terms of ease of implementation. The authors also commented that no direct mechanism was provided in PSALMS to cross-classify categorical demographic measures; the program was tedious to implement.

1 HESBRR (Health Examination Survey Variance and Crosstabulation program developed by the National Center for Health Statistics)

TABLE 2
Comparison of Computation Time

Variance of	CPU time in seconds			
	SESUDAAN/ RATIOEST	OSIRIS/ PSALMS	SUPER CARP	HESBRR [1/]
Means				
n= 1,193	2.51	6.27	39.28	8.21
n= 8,507	10.09	39.72	250.74	37.01
n= 40,320	42.95	184.62	--a	168.11
Totals				
n= 1,193	2.04	6.27	19.20	8.21
n= 8,507	7.31	39.72	117.47	37.01
n= 40,320	29.82	184.62	546.21	168.11
Ratios				
n= 1,193	5.69	15.18	47.74	37.73
n= 8,507	19.92	97.85	295.13	183.70
n= 40,320	79.97	456.62	--a	850.42

n indicates sample size   [a] Exceeded 15 minutes of CPU time

1/ HESBRR (Health Examination Survey Variance and Crosstabulation program developed by the National Center for Health Statistics)

As can be observed, the SESUDAAN/RATIOEST procedure was consistently superior in terms of computational efficiency over all specifications of data base and type of statistics used. Computational accuracy of SESUDAAN and SUPER CARP (both use TSL) was demonstrated earlier (Francis and Sedrank 1979). The variance estimates obtained from HESBRR were observed to be equal to hand calculated values by NCHS. A comparison of variance estimates derived by using HESBRR that uses BRR method with those obtained by TSL generally confirmed their computational convergence for large samples.

3. VARIANCE ESTIMATION FOR SIPP

First we note the key features of the SIPP design (documented fully in Nelson, McMillen and Kasprzyk 1984) before discussing how users can compute variances.

The SIPP has a stratified multi-stage design. In the 1984 Panel, there are 174 strata; 45 self representing (SR) and 129 nonself representing (NSR) strata. One PSU per NSR stratum is selected without replacement with probability proportionate to size. The ultimate sampling units are clusters of generally 2 housing units selected systematically from the 1970 decennial census address list. The sample is updated to reflect new construction with clusters of 4 housing units. Thus, the total sampling variance has three components; between and within PSU variances from NSR strata, and the within PSU variance from SR strata.

The SIPP is a rotating panel survey with a new panel begun each year that is interviewed at 4-month intervals for 2 1/2 years. The yearly panels beginning with the 1985 panel are selected from the 1980 census based design which has generally 2 PSUs per stratum whereas the 1984 Panel has 1 PSU per stratum. The 1980 census based design originally had 91 SR PSUs and 198 NSR PSUs (2 PSUs from each of 95 NSR strata and 1 PSU from the remaining 8 NSR strata). Before this design was ever implemented, the sample had to be reduced in FY 1985 for budgetary reasons. The sample reduction was achieved by dropping 54 NSR PSUs and about 3000 households from SR PSUs. Thus, the sample reduction in 1985 introduced another component of variance -- between strata within "super strata" that were formed for dropping sample PSUs. In FY 1986, additional long-term cost saving measures were instituted by further reducing the sample size by 15% for the 1985 and 1986 Panels. These sample reductions will, of course, increase sampling variances but unlike the 1985 sample cut, will not introduce another new component of variance. A detailed description of sample reductions for 1984, 1985 and 1986 Panels is given by Kasprzyk and Herriot (1986).

A "reasonably good" variance estimator should incorporate all components of variance and have a small, if any, bias. None of the three general purpose software packages, OSIRIS IV, SUDAAN or SUPER CARP will estimate the component of variance introduced by the 1985 sample cut. The replicate weighting method developed by Fay at the Bureau of the Census incorporates this component of variance.

SUPER CARP and SUDAAN are appropriate for stratified multi-stage sampling designs with simple random sampling without replacement at each stage but not for PPS sampling of PSUs. SUPER CARP allows computation of the Yates-Grundy Variance estimator (Cochran 1977, P.261) for the 2 PSUs per stratum design. This estimator is unbiased only for single-stage sampling. As Des Raj (1968, P.118), points out, this will result in a serious underestimation of the variance. A term must be added to represent fully within PSU variances. The SUDAAN/SESUDAAN'S variance formula for stratified two-stage sampling with simple random sampling at each stage can be manipulated to yield the appropriate variance formula for PPS sampling. This is indicated in Folsom (1985). However, this is not built into the package and would require additional work by a knowledgeable person and also computation of joint probabilities of selection of pairs of PSUs in each stratum. Another problem with this manipulation is that for confidentiality reasons, the SIPP Public use data file does not identify each PSU.

These complications can be avoided by using a replication method of variance computation. The OSIRIS IV/REPERR software uses the BRR method based on sampling of PSUs with replacement. This tends to overestimate the variance for a design like SIPP where PSUs are sampled without replacement. Thus, all three software packages have some deficiencies in meeting the objective of providing strictly design based estimates of variances for the complex design of the SIPP. But they will provide reasonably good estimates of variances for SIPP. Moreover, they provide substantial improvement over the traditional method based on the assumption of simple random sampling ignoring complex design features. Fay's replicate weighting method is geared to the specific sample design of the SIPP. Reweighting of each replicate independently to population control totals, hopefully, provides a more efficient variance estimator than the BRR method. Unfortunately, users will not be able to use this method to compute variances because replicate weights are not provided in public use files. The original PSU codes as well as replicate weights are withheld from public use files to prevent identification of small geographic areas, where disclosure of individual identities may be possible. However, users can compute variances by the BRR method using special codes provided in public use files. Each sample person in a SIPP public use file has been assigned a pseudo stratum code and a pseudo PSU (half-sample) code. Each pseudo stratum has two pseudo PSUs. Bye and Gallicchio (1989) have computed variances for characteristics of a subpopulation-- Old-Age, Survivors and Disability Insurance (OASDI) and Supplemental Security Income (SSI) program participants--by the BRR method from the 1984 SIPP public use file.

They also compared variances obtained by this method for 36 estimates of population totals with those obtained by the Bureau of the Census using Fay's method. Most of the items showed small differences in coefficients of variation. There are indications that variance estimates may have larger variances compared to Fay's variance estimates especially when the number of sample cases are small. This is not surprising since Fay's method is designed to increase the stability of variance estimates.

The overall results of this comparison however, show that the variances computed by the BRR method using pseudo strata and half-sample codes are comparable to the design based variances obtained by Fay's replicate weighting method. Users can compute variances by the BRR method without the help of any variance software package. A distinct advantage of the BRR method is that it uses a single variance formula for all statistics and thus facilitates computation of variances for complex statistics like socioeconomic indices for which derivation of variance formulas may not be easy. As mentioned in the section 1.5, empirical results suggest that the BRR method performs better than the TSL or JRR method in estimating confidence intervals for parameters. These advantages, in addition to the ease of computation, lead us to recommend that users compute variances directly when necessary, by the BRR method using pseudo stratum and half-sample codes provided in the public use file.

The minimum number of half-sample replicates required for a fully balanced orthogonal set is the smallest multiple of 4 which is greater than the number of strata in the sample design. Thus, the number of replicates, required is at most L+4 where L is the number of pseudo strata in a public use file. Let R denote the number of replicates and let $\hat{\theta}_i$ (i=1,2,...R) and $\hat{\theta}$ denote the estimate of the parameters θ as computed from the ith replicate and from the full sample respectively.

Then, the variance of $\hat{\theta}$ is estimated by

$$\hat{V}(\hat{\theta}) = \sum_{i=1}^{R} (\hat{\theta}_i - \hat{\theta})^2/R.$$

Note that when R replicates represent a fully balanced orthgonal set, $\hat{\theta}$ is equal to $\bar{\theta} = \sum_{i=1}^{R} \hat{\theta}_i/R$, for strictly linear statistics. But for nonlinear statistics like ratio, quantile, correlation coefficients etc $\hat{\theta}$ and $\bar{\theta}$ are not equal. They should be quite close, however, in most situations for moderate to large samples. Large differences among $\hat{\theta}$ and $\bar{\theta}$ should indicate that computational errors may have occurred.

The 1984 SIPP public file has 71 pseudo strata and therefore requires 72 half-sample replicates for full orthogonal balance as has been used by Bye and Gallicchio (1989). The SIPP public use files for 1985 and following years have 72 pseudo strata. Users should use 76 half-sample replicates to compute variances. A 76x76 orthogonal matrix (called Hadamard matrix) is given in Placket and Burman (1946) and Wolter (1985). However, users can generate an orthogonal matrix following the procedure described below:

Generating an Orthogonal Matrix

The following describes how to generate an orthogonal matrix which is a Hadamard matrix. The description is for a n x n matrix where n = 4t = p + 1, p being an odd prime. The matrix will be made up of +1 and -1 entries.

The first step in generating the matrix is to create the first row and the first column of the matrix by setting each entry to +1 so that for matrix m, m(1,i) = +1 and m(i,1) = +1 for i = 1 to n.

The next step is to generate the second row. First, letting p = n - 1 and q = n - 2 find the set S of perfect squares (mod p). To do this, calculate the elements s_h of S as follows:

$$s_h = \text{the remainder of } \frac{h^2}{p} \quad \text{for h = 1 to q}$$

Set m(2,2) = -1. For j = 3 to n, let k = j - 2. The remaining columns in the second row can be assigned as follows:

$$m(2,j)=\begin{cases}+1 & if\ k\varepsilon\ S\\ -1 & if\ k\notin S\end{cases}$$

The last step is to generate the remaining rows. For i = 3 to n and j = 2 to n, assign each column in each row as follows:

$$m(i,j)=\begin{cases}m(i-1,n) & if\ j = 2\\ m(i-1,j-1) & if\ j \neq 2\end{cases}$$

The final result will be an n x n Hadamard matrix. A 12 x 12 Hadamard matrix is shown below.

12 x 12 Hadamard Matrix

1	1	1	1	1	1	1	1	1	1	1	1
1	-1	1	-1	1	1	1	-1	-1	-1	1	-1
1	-1	-1	1	-1	1	1	1	-1	-1	-1	1
1	1	-1	-1	1	-1	1	1	1	-1	-1	-1
1	-1	1	-1	-1	1	-1	1	1	1	-1	-1
1	-1	-1	1	-1	-1	1	-1	1	1	1	-1
1	-1	-1	-1	1	-1	-1	1	-1	1	1	1
1	1	-1	-1	-1	1	-1	-1	1	-1	1	1
1	1	1	-1	-1	-1	1	-1	-1	1	-1	1
1	1	1	1	-1	-1	-1	1	-1	-1	1	-1
1	-1	1	1	1	-1	-1	-1	1	-1	-1	1
1	1	-1	1	1	1	-1	-1	-1	1	-1	-1

Note that columns of the matrix can be associated with pseudo strata and rows with replication. An entry of +1 in the *i*th row and *j*th column (i,j) cell signifies that the half-sample 1 is part of the *i*th replicate, while an entry of -1 signifies that the half-sample 2 is part of the i-th replicate.

The sample records thus assigned to the *i*th replicate will have to be multiplied by a factor 2 to compute the estimate $\hat{\theta}_i$. This is equivalent to giving a replicate weight 2 to all records assigned to the *i*th replicate and a weight 0 to the remaining records. This can be done by generating replicate factors (2.0) as described below.

Generating replicate factors

The following describes how to generate replicate factors for a file using the Hadamard matrix.

In order to create replicate factors for each record on the file, the file will have a pseudostratum code. Each pseudostratum is divided into two pseudo PSUs or half-samples. The half-sample code on each record is usually a 1 or 2, although other possibilities may exist. The following details will assume that the code is represented as either a 1 or 2 on each record.

At this time a n x n Hadamard matrix should have been generated, where n is greater than the number of pseudostrata but equal to the number of replicates to be formed. Associate each pseudostratum code with a column on the matrix. For each record, find the column associated with the pseudostratum code. Then create replicate factors form the column of the matrix. To do this, let the following variables be defined as:

R :	total number of replicates where R=n
s :	half-sample code for a given record
m :	Hadamard matrix
c :	column of the matrix associated with the pseudostratum code for a given record
$\{f_i\}$:	replicate factors for a given record

To generate all replicate factors use the following procedure.

If s=1 then

$$f_i = \begin{cases} 2.0 & \text{if } m(i,c) = +1 \\ 0.0 & \text{if } m(i,c) = -1 \end{cases} \qquad \text{for i=1 to R}$$

If s=2 then

$$f_i = \begin{cases} 0.0 & \text{if } m(i,c) = +1 \\ 2.0 & \text{if } m(i,c) = -1 \end{cases} \qquad \text{for i=1 to R}$$

From this procedure R replicate factors are formed for a given record. Continue this procedure until all records on the file have replicate factors.

4. VARIANCE GENERALIZATION

The SIPP reports publish standard errors of many summary statistics. Because of costs, sampling errors are usually computed only for a representative subset of statistics. Generalized variances are obtained for many characteristics by fitting statistical models. The appendix 'Source and

Reliability of Estimates' of a report provides standard errors related to the size of an estimate for key characteristics so that users can obtain standard errors by linear interpolation. Generalized parameters are also provided for direct computation of standard errors for some characteristics. Computations are illustrated in the appendix. The main purpose of the appendix is to enable readers to ascertain the reliability of estimates and make statistical inferences. Researchers can compute standard errors for studies involving general population characters.

In studies confined to small domains of the population, e.g. OASDI and SSI program participants, Hispanics etc., generalized parameters may not be appropriate, and direct computation of variances by the BRR would be useful. In studies involving a very large number of statistics, users may compute variances directly for a representative subset of statistics and fit a suitable linear model to generalize variances.

The model

$$V^2 = \alpha + \beta/x$$

where

α and β are coefficients to be estimated
x is the estimated population total

and V^2 is the estimated relative variance of x, i.e.

$$V^2 = Var(x)/x^2$$

has been used by the Census Bureau for CPS, SIPP and other surveys. Users can use this model to obtain generalized variances.

5. RECOMMENDATIONS FOR VARIANCE ESTIMATION

1. In studies involving general population characteristics, users can compute variances by using the generalized parameters provided in SIPP reports.

2. In studies involving small domains of the population, users should directly compute variances by the BRR method.

3. Variances computed on a small number of sample cases may not be reliable and therefore, caution should be exercised in making statistical inferences based on such variance estimates.

REFERENCES

Anderson, C., Forsman, G. and Wretman, J. (1987). Estimating the Variance of a Complex Statistic: A Monte Carlo Study of Some Approximate Techniques. Journal of Official Statistics, 3, 251-265.

Arvensen, J.N. (1969). Jackknifing U-statistics. Annals of Mathematical Statistics, 40, 2076-2100.

Bean, J.A. (1975). Distribution and Properties of Variance Estimators for Complex Multi-Stage Probability Samples. Vital and Health Statistics, Series 2 No. 65, U.S. Govt. Printing Office, Washington D.C.

Brillinger, D.R. (1964). The Asymptotic Behavior of Tukey's General Method of Setting Approximate Confidence Limits (The Jackknife) when Applied to Maximum Likelihood Estimates. Review of International statistical Institute 32, 202-206.

Bye, B.V. and Gallicchio, S.J. (1989) Two Notes on Sampling Variance Estimates from the 1984 SIPP Public - use Files. SIPP Working papers No. 8902 April 1989. Bureau of the Census, Washington, D.C.

Campbell, C. and Myer, M. (1978). Some Properties of T Confidence Intervals for Survey Data. Proceedings of American Statistical Association, Survey Research Methods Section, 437-442.

Chakrabarty, R.P. and Rao, J.N.K. (1967). The Bias and Stability of Jackknife Variance Estimator in Ratio Estimation. Proceedings of American Statistical Association, Social Statistics Section. 326-331.

Chakrabarty, R.P. (1989). Multivariate Analysis by Users of SIPP Micro - Data Files. SIPP Working Paper Series No. 8915. U.S. Bureau of the Census, Washington, D.C.

Cochran, W.G. (1977). Sampling Techniques. John Wiley & Sons, New York.

Cohen,S.B. and Gridley, G. (1981). Present Limitations in the Availability of Statistical Packages for the Analysis of Complex Survey Data. Proceedings of American Statistical Association, Statistical Computing Section. 20-24.

Cohen, S.B., Burt, V.L. and Jones, G.K. (1986). Efficiencies in Variance Estimation for Complex Survey Data. The American Statistician, 40, 157-164.

Deming, W.E. (1956). Simplifications of Sampling Design through Replication with Equal Probabilities and without Stages. Journal of the American Statistical Association, 51, 24-53.

Deming, W.E. (1960). On Sample Design in Business Research. John Wiley & Sons, New York.

Diaconis, P. and Efron, B. (1983). Computer-Intensive Methods in Statistics. Scientific American, 248, 116-130.

Dippo, C.S. and Wolter, K.M. (1984). A Comparison of Variance Estimators Using the Taylor Series Approximation. Bureau of the Census, SRD Research Report Number: Census/SRD/RR-84-21.

Dippo, C.S., Fay, R.E. and Morganstein, D.M. (1984). Computing Variances from Complex Samples with Replicate Weights. Proceedings of the American Statistical Association, Survey Research Methods Section.

Draper, N.R. and Smith, H. (1966). Applied Regression Analysis. John Wiley & Sons, New York.

Efron, B. (1982). The Jackknife, the Bootstrap and Other Resampling Plans. Society for Industrial and Applied Mathematics, Philadelphia, PA.

Fay, R.E. (1984). Some Properties of Estimates of Variance Based on Replication Methods. Proceedings of the American Statistical Association, Survey Research Methods Section.

Fay, R.E. (1989). Theory and Application of Replicate Weighting for Variance Calculations. Proceedings of the American Statistical Association. Survey Research Methods Section.

Folsom, R.E. (1985). Unbiased Variance Estimation for the Two-Stage Horvitz-Thompson Estimator by Tricking SESUDAAN. Unpublished Note.

Francis, I. and Sedransk, J. (1979). A Comparison of Software for Processing and Analyzing Survey Data. Bulletin of International Statistical Institute, 48, 1-31.

Frankel, M.R. (1971). Inference from Survey Samples. Institute for Social Research. Ann Arbor, Michigan.

Fuller, W.A. (1975). Regression Analysis for Sample Surveys. Sankhya, series C, 37, 117-32.

Holt, M.M. (1977). SURREGR: Standard Errors of Regression Coefficients from Sample Survey Data. Research Triangle Institute. N.C.

Kasprzyk, D. and Herriot, R. (1986). Some Aspects of the Survey of Income and Program Participation. SIPP Working Paper Series No. 8601, April. U.S. Bureau of the Census.

Keyfitz, N. (1957). Estimates of Sampling Variance Where Two Units are Selected from Each Stratum. Journal of the American Statistical Association 52, 503-510.

Kendall, M.G. and Stuart, A. (1963). Advanced Theory of Statistics, Vol. I. Charles Griffin & Company, London.

Kish, L. (1968). Standard Errors for Indexes from Complex Samples. Journal of the American Statistical Association, 63, 512-529.

Kish, L. and Frankel, M.R. (1974). Inference from Complex Samples. Journal of the Royal Statistical Society, Series B, 36, 1-37.

Kovar, J., Rao, J.N.K. and Wu, C.F.J. (1988). Bootstrap and Other Methods to Measure Errors in Survey Estimates. Canadian Journal of Statistics, 16. Supplement, 25-45.

Krewski, D. and Chakrabarty, R.P. (1981). On the Stability of the Jackknife Variance Estimator. Journal of Statistical Planning and Inference, 5, 71-79.

Krewski, D. and Rao, J.N.K. (1981). Inference from Stratified Samples: Properties of the Linearization, Jackknife and Balance Repeated Replication Methods, The Annals of Statistics 9, 1010-1019.

Lemeshow, S. and Levy, P. (1978). Estimating the Variance of Ratio Estimates in Complex Sample Surveys with two Primary Units per stratum-A Comparison of Balanced replication and Jackknife techniques. Journal of Statistical Computation and Simulation, 8, 191-205.

Mahalanobis, P.C. (1939). A Sample Survey of the Acreage Under Jute in Bungal. Sankhya, 4, 511-531.

Mahalanobis, P.C. (1944). On Large-scale Sample Surveys. Phil. Transactions, Royal Society. London, B231, 329-451.

McCarthy, P.J. (1966). Replication: An Approach to the Analysis of Data from Complex Surveys. National Center for Health Statistics, Washington D.C. Series 2, 14.

McCarthy, P.J. (1969). Pseudo replication: Half Samples. Review of International Statistical Institute, 37, 239-264.

Mulry, M. and Wolter, K.M. (1981). The Effect of Fisher's Z-transformation on Confidence Intervals of the Correlation Coefficient. Proceedings of the American Statistical Association, Survey Research Methods Section.

Nelson, D., McMillen, D. and Kasprzyk, D. (1984). An Overview of the Survey of Income and Program Participation. SIPP Working Paper Series No. 8401, September. U.S. Bureau of the Census.

Plackett, R.L. and Burman, P.J. (1946). The Design of Optimum Multi-factorial Experiments. Biometrika, 33, 305-25.

Quenouille, M.H. (1956). Notes on Bias in Estimation. Biometrika, 43, 353-360.

Raj, D. (1968). Sampling Theory. McGraw-Hill, New York.

Rao, J.N.K. (1968). Some Small-Sample Results in Ratio and Regression Estimation. Journal of the Indian Statistical Association, 6, 160-168.

Rao, J.N.K. (1969). Ratio and Regression Estimators. New Developments in Survey Sampling, N.L. Johnson and H. Smith, Jr. (eds.), John Wiley & Sons, New York, 213-234.

Rao, J.N.K. (1988). Variance Estimation in Sample Surveys. In Handbook of Statistics, vol. 6. (Eds. P.R. Krishnaiah and C.R. Rao). Amsterdam: Elsevier Science, 427-447.

Rao, J.N.K. (1993). Resampling Methods for Complex Surveys. Washington Statistical Society Short Course. March 1.

Rao, J.N.K. and Beagle, L.D. (1967). A Monte Carlo Study of Some Ratio Estimators, Sankhya. Series B, 36,47-56.

Rao, J.N.K. and Kuzik, R.A. (1974). Sampling Errors in Ratio Estimation. Sankhya. Series C, 36, 43-58.

Rao, J.N.K. and Wu, C.F.J. (1985). Inference From Stratified Samples: Second-Order Analysis of Three Methods for Nonlinear Statistics. Journal of the American Statistical Association 80, 620-630.

Rao, J.N.K. and Wu, C.F.J. (1988). Resampling Inference with Complex Survey Data. Journal of the American Statistical Association. 83, 231-241.

Rao, J.N.K., Wu, C.F.J. and Yue, K. (1992). Some Recent Work on Resampling Methods for Complex Surveys. Survey methodology. 18, 209-217.

Rust, K. (1985). Variance Estimation for Complex Estimators in Sample Surveys. Journal of the Official Statistics, 4, 381-397.

Shah, B.V., Holt, M.M. and Folsom, R.E. (1977). Inference about Regression Models From Sample Survey Data. Bulletin of International Statistical Institute, 47, 43-57.

Sukhatme, P.V., Sukhatme, B.V., Sukhatme, S. and Asok, C. (1984). Sampling Theory of Surveys with Applications. Iowa State University Press, Ames, Iowa.

Tepping, B. (1968). The Estimation of Variance in Complex Surveys. Proceedings of the American statistical Association, Social Statistics Section, 11-18.

Tukey, J.W. (1958). Bias and Confidence in Not-Quite Large samples. Annals of Mathematical Statistics 29, 614 (Abstract).

Wolter, K.M. (1985). Introduction to Variance Estimation. Springler-Verlag, New York.

Woodruff, R.S. (1971). A Simple method for Approximating the Variance of A Complicated Estimate. Journal of the American Statistical Association, 66, 411-414.

Woodruff, R.S. and Causey, B.D. (1976). Computerized Method for Approximating the Variance of a Complicated Estimate. Journal of the American Statistical association, 71, 315-321.

APPENDIX

A List of Variance Estimation Programs

	Package Name	Developer	Affiliated Institution
1.	OSIRIS IV	L. Kish, et al	Survey Research Center University of Michigan Ann Arbor, MI 48106
2.	SUDDAN	B.V. Shah	Research Triangle P.O. Box 12194 Research Triangle Park, NC 22709
3.	SUPER CARP	M. Hidiroglou	Iowa State University W. FullerDepartment of Statistics R. HickmanAmes, Iowa 50010
4.	CLUSTERS	V. Verma	World Fertility Survey M. PearceInternational Statistical Institute Voorburg, The Netherlands
5.	HESBRR	G. K. Jones	Center for Health Statistics 3700 Eastwest Highway Hyattsville, MD 20782
6.	NASSTIM, NASSTVAR	D. Morgenstein	Westat, Inc. 1650 Research Blvd. Rockville, MD 20850
7.	PASS	D. Thompson	Social Security Administration R. Finch, Jr.6401 Security Blvd. Baltimore, MD 21235
8.	RGSP	F. Yates	Rothamsted Experimental Station Harpenden, Hertfordshire England AL52JQ
9.	SPLITHALVES	J.R. Pryor	Australian Bureau of Statistics P.O. Box 10, Belconnen A.C.T. 2616, Australia

10. CAUSEY	B. Causey R. Woodruff	U.S. Bureau of the Census Statistical Research Division Washington, D.C. 20233
11. TREES	D. Bellhouse	Department of Statistics Univ. of Western Ontario London ON N6A 5B9 Canada
12. VTAB, SMED 83	J. Wretmen	Statistics, Sweden Stockholm S-1115, Sweden
13. U-SP	G.B. Wetherill	Mathematical Institute et alUniv.of Kent at Canterbury Kent, England CT27NF
14. FINSYS-2	W.E. Frayer	School of Forestry Colorado State Univ. Fort Collins, Colorado