Survey Methodology, June 1988
Vol. 14, No. 1, pp. 59-69
Statistics Canada

Personal Computer Variance Software for Complex Surveys

DAN SCHNELL, WILLIAM J. KENNEDY, GARY SULLIVAN, HEON JIN PARK and WAYNE A. FULLER[1]

ABSTRACT

A personal computer program for variance estimation with large scale surveys is described. The program, called PC CARP, will compute estimates and estimated variances for totals, ratios, means, quantiles, and regression coefficients.

KEY WORDS: Survey sampling; Variance estimation; Survey software.

1. INTRODUCTION

The analysis of survey data typically involves a large number of observations and relatively complex variance calculations. Recent developments in personal computers have made possible the use of such computers to process data from complex surveys. We describe a personal computer program for survey data analysis prepared at Iowa State University.

The project to develop statistical software for variance estimation on the personal computer was a joint undertaking between Iowa State University and the International Statistical Programs Center of the U.S. Census Bureau. The objective of the Census Bureau was to provide developing countries with software that can be used locally to process survey data collected locally. The Iowa State University project on variance estimation was part of a larger Census Bureau undertaking that included the development of software for survey management, data editing and tabulation.

Beginning in the early 1970's, based on the work of Hidiroglou (1974) and Fuller (1975), a program was developed at Iowa State University for the computation of regression coefficients and the estimated covariance matrix of the coefficients for survey data. The program, called SUPER CARP, was later expanded to include total estimation, ratio estimation, subpopulation statistics, two-way tables and two stage samples. The last revision of SUPER CARP took place in 1980. SUPER CARP furnished the starting point for software development on the personal computer. Because of its ancestry, the personal computer program was called PC CARP.

2. PROGRAM CAPABILITY

PC CARP was designed for the IBM PC, IBM PC/XT, IBM PC/AT and compatible machines. At least 410K bytes of memory and a math coprocessor are required.

PC CARP is capable of handling both large and small data sets with equal ease and efficiency. The program sets no limit on the number of strata or clusters that can appear in a data set and can accept up to 50 input variables at a time. The program accepts disk data files in either fixed or free format.

The program can be used to compute variances for one or two stage samples with finite population correction terms included. For samples with more than two stages, finite population

[1] Dan Schnell, Centers for Disease Control, 1600 Clifton Road, NE, Atlanta, Georgia 30333. William J. Kennedy, Gary Sullivan, Heon Jin Park and Wayne A. Fuller, Department of Statistics, Iowa State University, Ames, Iowa 50011, United States.

corrections are only available at two levels. For two-stage samples, the program computes within cluster sampling rates from the stratum sampling rates and the individual observation weights.

Typically, each observation in the data file will contain stratum identification, cluster (primary sampling unit) identification, and a weight where the weight is the inverse of the selection probability. The user may or may not elect to enter first stage sampling rates. For simple designs, such as simple random sampling, not all of this information is required. In such cases reduced data input is possible.

If stratification is present, the program requires that all observations belonging to the same stratum be grouped together. If clustering is present, all observations belonging to the same cluster must be grouped together.

Table 1 contains a description of the types of statistics available to the user of PC CARP. In addition to the items of Table 1, supplements are available for estimation of the logistic function and for post stratified samples. These supplements are discussed in Section 4.4 and Section 4.5. An "X" in the column headed "Cov. matrix" means that the covariance matrix of a vector of estimates of the type listed on the left can be obtained. The standard error is computed for all statistics, but the covariance matrix of a vector is available for only a restricted set. Also, the coefficient of variation is computed for many statistics. The design effect, denoted by DEFF, is available as an option for many of the statistics. See Kish (1965) for a description of the design effect.

Table 1

Analysis Capabilities of PC CARP

Analysis	Coeff. var.	Cov. marix	Design effect	Comments
Population Analyses				
Total Estimation	X	X	X	50 variables maximum
Ratio Estimation	X	X	X	50 variables maximum without covariances 15 with covariances
Difference of Ratios			X	15 variables maximum
Stratum Analyses				
Totals	X		X	50 variables maximum
Means	X		X	50 variables maximum
Proportions	X		X	50 variables maximum
Subpopulation Analyses				
Totals	X		X	Crossed classif.
Means	X		X	Multiple variables
Proportions	X		X	Crossed classif. Multiple variables
Ratios	X		X	Crossed classif. Multiple variables
Other Analyses				
Two-way Table		X		50 cells maximum proportionality test
Regression		X		50 variables maximum Multiple d.f. tests Y-hat, residuals
Univariate			X	Multiple variables, empirical CDF, quantiles

The population (Total, Ratio and Difference of Ratios) analyses and stratum analyses are performed in a straightforward manner. Some details pertaining to Subpopulation Analyses, the Two-Way Table, Regression Analysis, and Univariate Analysis are presented in Section 4.

The subpopulation analyses give the user the option of crossing classification variables. This allows the user to create new classification structures from two or more input classification variables. For example, suppose the input data includes the classification variables age, sex and education with six, two and five levels, respectively. Then, by crossing age with sex with education, a new classification structure with 60 levels is produced. The user may obtain estimates for any number of dependent variables under this classification structure.

The Two-way Table analysis is defined by two classification variables and a dependent variable. More than one dependent variable can be specified for a pair of classification variables. Tables of cell totals, of proportions based on row totals, of proportions based on column totals, and of proportions based on the grand total are computed for each dependent variable. Standard errors are computed for all estimators and a test statistic for the hypothesis of proportionality is output. The test statistic is based on a Satterthwaite approximation to the distribution of the Pearson chi-square statistic. Also see Rao and Scott (1984).

The weighted least squares regression analysis computes coefficient estimates, and an estimated variance-covariance matrix which takes into account the sample design. These calculations are given in Fuller (1975) and outlined in Hidiroglou *et al.* (1980). Multiple degrees of freedom F-tests for sets of coefficients and the usual t-statistics are available. The user also has the option of obtaining residuals and predicted values.

The Univariate analysis provides statistics that describe the distribution of a variable. The user specifies the variable of interest and identifies a subpopulation by specifying a category of a classification variable. Thus, the user might elect to obtain statistics for the personal income of individuals in the professional category of the occupation classification. Estimates of the subpopulation mean, variance, distribution function, quantiles and interquartile range are produced.

3. PROGRAM DETAILS

PC CARP is written almost entirely in FORTRAN, the most widely known scientific programming language, and the IBM Professional FORTRAN compiler was selected for the project. A small portion of the code — some sections of the user interface — is written in IBM Assembly language.

Two concerns at the program development stage were to provide a friendly user interface and to minimize the number of passes through the data. The interface was made user friendly by implementing an interactive, screen oriented response system. A single pass algorithm for variance estimation of simple statistics minimized the amount of reading from data files. Most estimators and their variances are obtained in a single pass through the data.

Estimators can be computed for the total population, for each stratum, or for specified subpopulations. For the most part, the estimators are functions of weighted sample totals. For example, to compute the estimators of the ratios, $R_1 = Y_1/X_1$ and $R_2 = Y_2/X_2$, one accumulates totals for Y_1, X_1 Y_2, and X_2. If the estimate is for the entire population, these totals are accumulated in one pass through the data. Totals for stratum estimates can be accumulated, combined if necessary, and output stratum by stratum. Since the data are grouped by strata, stratum totals can also be obtained in one pass for any number of strata. Subpopulation estimators may require more than one pass through the data if the number of categories defined by the classification structure is large. The Regression and Univariate analyses require two passes through the data.

The estimators, with the exception of totals, are nonlinear functions of weighted sample moments. It follows that a method appropriate for a nonlinear function must be used to estimate the variance of the approximate distribution of such estimators. See Wolter (1985) for a discussion of variance estimation for complex surveys. The Taylor method (method of statistical differentials) is the method of variance estimation used in PC CARP. Generally, the Taylor method has been shown to be equal to or superior to other variance estimation methods for the statistics, such as ratios, under consideration. See, for example, Frankel (1971). The Taylor variance of the ratio estimator is given in such standard texts as that of Cochran (1977) and the Taylor variance of a regression coefficient is given by Fuller (1975).

The value of the estimator and its estimated variance can, in most cases, be computed in the same pass. This is because the first order Taylor approximation to the variance can be expressed in terms of the variances of totals. For example, the first order Taylor approximation to $\hat{R} = \hat{Y}/\hat{X}$ is

$$\hat{R} \doteq R + X^{-1}(\hat{Y} - R\hat{X}),$$

where $R = Y/X$ is the ratio of the true totals. It follows that the estimated variance of a ratio $\hat{R} = \hat{Y}/\hat{X}$ can be computed from the estimated variance of the totals of Y, X, and $(Y - X)$. Similarly, the estimated covariance matrix for $\hat{R}_1 = \hat{Y}_1/\hat{X}_1$ and $\hat{R}_2 = \hat{Y}_2/\hat{X}_2$, can be computed from the estimated variances of the totals of the ten quantities $Y_1, X_1, (Y_1 - X_1), Y_2, X_2, (Y_2 - X_2), (Y_1 - Y_2), (Y_1 - X_2), (Y_2 - X_1)$, and $(X_1 - X_2)$.

The algorithm used for the calculation of the weighted mean and weighted sums of squares and cross products matrices is described in Herraman (1968). For sample values $\{X_i\}$ and corresponding weights $\{W_i\}$, the sequence of weighted means, $\bar{X}_K$, and weighted corrected sum of squares, S_K, is computed as

$$\bar{X}_K = \bar{X}_{K-1} + a_K d_K \quad \text{and} \quad S_K = S_{K-1} + D_K - D_K a_K,$$

where $d_k = X_K - \bar{X}_{K-1}$, $a_K = W_K(\Sigma_{i=1}^{K} W_i)^{-1}$, and $D_K = d_k^2 W_K$.

Up to three different variance quantities can be accumulated concurrently for any given estimator. These are the first stage variance component, the optional second stage variance component and the optional simple random sampling variance used in the computation of the design effect. Computing all variance quantities in a single pass through the data requires a large amount of array space. However, when working with large samples, the elimination of entire passes through the data out-weighs the use of additional memory.

The program routinely performs checks to avoid computational errors such as division by zero. For example, if the user enters a data set with only one cluster in a stratum, the program will assign zero variance to the stratum, complete the calculations, and print an error message identifying the stratum with a single cluster.

The error handling system was constructed to avoid program termination caused by user misspecifications that could be easily corrected. Checks for omitted responses, improper file names and invalid analysis variable specifications are included in the program. If such an error is detected, PC CARP permits the user to re-enter information or to exit the program.

Program accuracy was assessed by constructing examples and comparing results with those obtained using the mainframe program SUPER CARP. The data set of Longley (1967) was used to evaluate the accuracy of the regression program. Additional checks were made using PROC MATRIX of the SAS package. See Barr, *et al.* (1979). PC CARP numerical accuracy was found to be at the same level as the mainframe packages. Internal consistency of PC CARP was also verified by computing equivalent estimators using different options, e.g., by computing a subpopulation mean using the subpopulation option and using the ratio option.

When information is needed by PC CARP, the user receives a full screen of short response questions along with detailed instructions. The first set of screens displayed to the user ask for information pertaining to data organization and location. "Help" and "Go Back" options are available at many places.

The second phase of program execution is Analysis Specification. In this phase the user chooses the type of analysis, options for that type of analysis, and the analysis variables. Any number of analyses can be performed using the data specified in phase one.

4. SPECIAL FEATURES

4.1 Two Way Table

As described in Section 2, this option automatically provides the user with four tables, where the entries are determined by the type of marginal control exercised in constructing the table.

We outline the procedure used to construct the table of cell proportions and the estimated covariance matrix of the proportions. Suppose the table has R rows and C columns and let $\hat{Y}_{rc}$ be the estimated total for the *rc-th* cell. Let $\hat{Y}$ be the RC-dimensional column vector of cell totals, created by listing the columns of totals one beneath the other beginning with the first column. Let

$$\hat{Y}_{..} = \sum_{r=1}^{R} \sum_{c=1}^{C} \hat{Y}_{rc},$$

$$\hat{P}_{rc} = \hat{Y}_{..}^{-1} \hat{Y}_{rc}$$

be the estimated population total and the estimated cell proportion for cell rc, respectively.

Let $\hat{P}$ be the RC-dimensional column vector, analogous to $\hat{Y}$, composed of the RC values $\hat{P}_{rc}$, arranged by column. The estimated covariance matrix for $\hat{P}$ is

$$\hat{V}_{PP} = \hat{Y}_{..}^{-2} [I_{RC} - (\hat{P} \otimes J'_{RC})] \hat{V}_{YY} [I_{RC} - (\hat{P} \otimes J'_{RC})]',$$

where $\hat{V}_{YY}$ is the estimated covariance matrix of the vector of cell totals $\hat{Y}$, I_{RC} is the identity matrix of dimension RC, and J_{RC} is an RC-dimensional column vector of ones.

The matrix $\hat{V}_{PP}$ is used to compute the test statistic for the hypothesis of proportionality. The null hypothesis for the test is the hypothesis that the interior entries in the population table are the products of the marginal proportions. See Rao and Scott (1984) for a discussion of tests for such hypotheses. The test in PC CARP is based on a Satterthwaite approximation to the distribution of the Pearson chi-square statistic constructed as if the proportions were multinomial proportions. The approximation is valid for any analysis variable.

4.2 Quantile Estimation

Among the statistics produced by the univariate option are estimates of quantiles and an estimator of the standard error of the quantiles. The first step in the computation of quantiles is the construction of an estimator of the cumulative distribution function. In a first pass through the data the range of observations, the sample mean, and the sample standard deviation are constructed. Also, the three largest observations and the three smallest observations are identified.

The estimated cumulative distribution function is defined by

$$\hat{F}_Y(x) = \left(\sum_{t=1}^{m} w_t Z_{St} \right)^{-1} \sum_{t=1}^{m} w_t Z_{St} I_Y(x),$$

where the summation is over the m elements in the sample, w_t is the sample weight, Z_{St} is an indicator function that is one if the observation is in the subpopulation of interest and zero otherwise, and $I_Y(x)$ is one if $Y < x$ and is zero otherwise. The range of the variable is divided into 100 intervals and the cumulative distribution function is estimated at the 101 values defined by this subdivision.

The covariance matrix for the estimated distribution function evaluated at 25 points, $j = 1, 5, \ldots, 96$, is estimated. The estimated standard errors are smoothed with a three point moving average and interpolation is used to obtain an estimated standard error for each of 101 points of the estimated distribution function. Linear interpolation is used to create an estimated distribution function that is monotone increasing. Using the smoothed standard errors, a monotone increasing upper bound and monotone increasing lower bound that form a pointwise 95% confidence interval for the distribution function are established. These bounds are then inverted to form 95% confidence interval for the quantiles. The interquartile range and its standard error are also estimated.

The quantile estimation is based on a theory that assumes the existence of an underlying superpopulation distribution function with a positive density. See Francisco (1987) for theoretical details and Park (1987) for computational aspects.

4.3 Regression Estimation

Estimates of the coefficients of a linear regression model are computed by the method of weighted least squares. Using the procedure given in Fuller (1975), an estimator of the covariance matrix of the coefficient vector is computed, taking into account the sample design.

The coefficient vector is estimated by

$$\hat{b} = (X'WX)^{-1}X'WY,$$

where X is the $n \times p$ matrix of independent variable values, $\hat{Y}$ is the n-dimensional vector of dependent variable values, W is a matrix with the observation weights on the diagonal and zeros elsewhere, and n is the total number of observations. The variance of $\hat{b}$ is estimated by

$$\hat{V}(\hat{b}) = (X'WX)^{-1}\hat{G}_W(X'WX)^{-1}.$$

The matrix $\hat{G}_W$ is

$$\hat{G}_W = C \sum_{i=1}^{L} h_i \sum_{j=1}^{n_i} (\hat{d}_{ij} - \bar{d}_{i.})(\hat{d}_{ij} - \bar{d}_{i.})',$$

where

$$\hat{d}_{ij} = \sum_{k=1}^{m_{ij}} X_{ijk} \hat{v}_{ijk} W_{ijk},$$

$$\hat{v}_{ijk} = Y_{ijk} - \hat{b}' X_{ijk},$$

$$\bar{d}_{i.} = n_i^{-1} \sum_{j=1}^{n_i} \hat{d}_{ij},$$

$h_i = (n_i - 1)^{-1} n_i$, $C = (n - p)^{-1}(n - 1)$, m_{ij} is the number of elements in cluster j of stratum i, n_i is the number of clusters in stratum i, n is the total number of observations, L is the number of strata, and p is the number of coefficients estimated. The variance estimator differs from the usual weighted least squares variance estimator in that the matrix $\hat{G}_W$ is used in place of $(X' WX)s^2$.

A multiple R-squared statistic is computed for models with an intercept. An F-test for the overall regression is always computed and an option for testing subsets of coefficients is provided.

4.4 Logistic Regression

Estimates of the multivariate logistic model are obtained with this option. The algorithms for logistic regression were developed after the initial version of PC CARP was completed. Because the mean function for the logistic model is nonlinear in the parameters, the estimates are computed using an iterative weighted least squares algorithm. The variances of the estimates are computed by the extension to nonlinear estimation of the procedures given in Fuller (1975). See also Binder (1983). The basic operation of the Logistic Regression option is the same as that of the Regression option. For example, independent and dependent variables are specified in the same way.

4.5 Post Stratification

After completion of the original PC CARP program a supplement for post stratification was developed for many of the estimators. The post stratification is assumed to be that in which the weights have been adjusted to produce estimates for certain categories that match known population totals. This type of post stratification is called gamma post stratification by Fuller and Sullivan (1987).

The program computes the variance of the post stratification estimator based on a representation in which the estimator is expressed as a sum of ratio estimators.

4.6 Stratum Collapse

For purposes of variance computation, the user may use the collapse option to eliminate one cluster strata. If this option is chosen, every one-cluster stratum is grouped with the immediately following stratum in the data set. The stratum and cluster identification of the involved records are changed to reflect the new stratification. If stratum sampling rates are present, new rates are defined by

$$f_i^* = (n_i f_i^{-1} + n_{i+1} f_{i+1}^{-1})^{-1}(n_i + n_{i+1}),$$

where stratum i, with $n_i = 1$, has been combined with stratum $i + 1$. These new rates are also saved in an auxiliary rate file for possible future use. Different orderings of the strata will produce different collapsed data sets and different collapsed stratum rates. The program requires an additional pass through the data when either the collapse or the two-stage option is selected.

4.7 Hot Deck Imputation

PC CARP requires a complete data set for analysis. Many practitioners will write a special program, or use one of the readily available PC programs to edit their data and to impute for missing values.

For those desiring it, a hot deck imputation program, called PRE CARP, is provided with PC CARP. The hot deck operation replaces a missing value with the value for the same item from the record immediately preceeding the missing record in the data file. PRE CARP permits the user to specify a classification variable, containing up to ten categories, such that the missing value is replaced by the preceeding record in the same category. PRE CARP will also create an indicator variable for each variable with missing values. This indicator variable can then be used with the subpopulation option to compute means based on the original observations.

5. EXAMPLES

In this section, several analyses are performed with a constructed data set and run times are presented. The purpose of the test runs is not to examine all possible combinations of factors influencing processing time, but rather to give an idea of the time required to run some of the available program analyses.

The test data were constructed from a subset of the second National Health and Nutrition Examination Survey (NHANES II). The test data set has 2400 observations which are divided into 32 strata. Each stratum has two primary sampling units and the primary sampling units are of varying sizes. Each observation also has a non-zero sampling weight.

Figure 1. Output for Example C, Mean Age by Sex and Race Combinations

```
                                  Subpopulation Means
Dependent variable is Age
      Category               Estimate           S.E.          C.V.          DEFF
   Sex  =  1.0000            Race  =  1.0000
                             3.06811D+01  6.19678D-01   2.0197D-02   1.3967D+00
   Sex  =  1.0000            Race  =  2.0000
                             3.13016D+01  7.88580D-01   2.5193D-02   9.5384D-01
   Sex  =  1.0000            Race  =  3.0000
                             3.41579D+01  2.24111D+00   6.5610D-02   2.0965D+00
   Sex  =  2.0000            Race  =  1.0000
                             1.33742D+01  3.18904D-01   2.3845D-02   1.2588D+00
   Sex  =  2.0000            Race  =  2.0000
                             *********
   Sex  =  2.0000            Race  =  3.0000
                             1.71957D+01  9.53816D-01   5.5468D-02   1.1389D+00
```

Figure 2. Univariate Output for Nonfarm Households for Example D

UNIVARIATE 1

Classification variable is Farm and its level is 1

Number of Sample Elements in Subpopulation = 111

Dependent variable is Age

Subpopulation Variance = 4.25645D + 02

Subpopulation C.V. = 7.14101D-01

Subpopulation Mean

Estimate	S.E.	C.V.	DEFF
2.8891089D + 01	2.2543875D + 00	7.80305D-02	9.86336D-01

Extreme Values of Sample Elements in Subpopulation

Smallest Values	Number of Observed Values	Stratum	First Observation ID Cluster	Weight
1.000D + 00	1	32	1	3.000D + 00
2.000D + 00	1	15	1	2.000D + 00
3.000D + 00	3	10	2	3.000D + 00

Largest Values	Number of Observed Values	Stratum	First Observation ID Cluster	Weight
7.400D + 01	2	29	1	3.000D + 00
7.100D + 01	2	7	1	2.000D + 00
7.000D + 01	4	7	1	2.000D + 00

Quantiles

	Estimate	S.E.	95% Confidence Interval
0.01	2.2690811D + 00	7.3167585D-01	(8.05729D-01, 3.73243D + 00)
0.05	4.2364814D + 00	1.1977759D + 00	(3.34942D + 00, 8.14053D + 00)
0.10	7.7750203D + 00	1.3563759D + 00	(5.36691D + 00, 1.07924D + 01)
0.25	1.3652930D + 01	1.4225576D + 00	(9.99238D + 00, 1.56826D + 01)
0.50	1.9449315D + 01	2.2740912D + 00	(1.57192D + 01, 2.48156D + 01)
0.75	4.5698071D + 01	4.7577709D + 00	(3.58936D + 01, 5.49247D + 01)
0.90	6.2787426D + 01	2.3472775D + 00	(5.51417D + 01, 6.45308D + 01)
0.95	6.5923423D + 01	1.2228344D + 00	(6.43837D + 01, 6.92750D + 01)
0.99	7.1714993D + 01	1.1425033D + 00	(7.05401D + 01, 7.40000D + 01)

Interquartile Range

Estimate	S.E.
3.2045141D + 01	4.2434890D + 00

The variables in the data set are:

1. Sex 1 = male, 2 = female
2. Race 1 = white, 2 = black, 3 = other
3. Farm 1 = non-farm household, 2 = farm household
4. Income Household income in thousands of dollars
5. Age Age in years.

A variable whose value is one for every observation (intercept variable) was created by the program. The analyses performed were:

A. Mean income for the sampled population
B. Mean income by stratum
C. Mean age for the two way classification of sex and race
D. Sample distribution functions of age for farm and non-farm groups.

Analysis A, estimating mean income, was performed using the Ratio option with Income as the numerator variable and the intercept variable in the denominator. The estimates of mean income by stratum, Analysis B, were computed directly with the Stratum Means option. Analysis C was performed with the Subpopulation Means option by crossing the classification variables Sex and Race and specifying Age as the dependent variable. The output from this analysis is given in Figure 1. The symbols "******" under the classification "Sex = 2 Race = 2" indicates that there were no observations falling into that classification category. The values of the design effects underscore the importance of taking into account the sampling design in the computation of estimated variances. For example, the design effect for the estimate with Sex = 1 and Race = 3 is approximately two. This means that the estimated variance of the sample mean for a simple random sample is one half of the variance estimate for the stratified cluster sampling plan. Characteristics of the distribution of Age for each of the two levels of the variable "Farm" were estimated using the Univariate option. The portion of the output for this example that pertains to nonfarm households is given in Figure 2. All the variances and standard error estimates given in this output take into account the sampling design.

The run times (in seconds) for analyses A, B, C and D for the 2,400 observations were 70, 135, 120 and 360, respectively. The runs were made on an IBM PC AT with the data stored on the hard disk and read in free format. Stratum sampling rates were not entered into the program. Output was routed to the monitor and to a disk file. Design effects for the estimates were requested in all of the analyses. The first three analysis require only one pass through the data for each analysis. More statistics are computed for analyses B and C than for analysis A. Analysis D requires 4 passes through the data, two passes for each univariate analysis.

ACKNOWLEDGEMENTS

In addition to the U.S. Census Bureau, support for PC CARP was provided by the Soil Conservation Service, U.S. Department of Agriculture. We thank the referees and the editors for useful comments. Persons in developing countries who are interested in the PC CARP program should contact: PC CARP, International Statistical Programs Center, U.S. Bureau of the Census, Washington, D.C., USA. 20233. Other interested persons should contact: PC CARP, Statistical Laboratory, Iowa State University, Ames, Iowa, USA 50011. Copies have been made available to institutions in more than 20 countries.

REFERENCES

BARR, A.J., GOODNIGHT, J.H., SALL, J.P., BLAIR, W.H., and CHILKO, D.M. (1979). SAS User's Guide. SAS Institute, Raleigh, North Carolina.

BINDER, D.A. (1983). On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review*, 51, 279-292.

COCHRAN, W.G. (1977). *Sampling Techniques*. New York: John Wiley.

DISKIN, B.A. (1985). Microcomputers in developing country statistical offices: current use and a look to the future. Proceedings of the 45th Session of the International Statistical Institute, Amsterdam.

FRANCISCO, C.A. (1987). Estimation of quantiles and the interquartile range in complex surveys. Unpublished Ph.D. dissertation. Iowa State University, Ames, Iowa.

FRANKEL, M.R. (1971). *Inference from Survey Samples*. Institute of Social Research, University of Michigan, Ann Arbor.

FULLER, W.A. (1975). Regression analysis for sample survey. Sankhyā C 37, 117-132.

FULLER, W.A., and SULLIVAN, G. (1987). Gamma Post Stratification. Report to the U.S. Bureau of the Census. Dept. of Statistics, Iowa State University, Ames, Iowa.

HERRAMAN, C. (1968). Sums of squares and products matrix. *Applied Statistics*, 17, 289-292.

HIDIROGLOU, M.A. (1974). Estimation of regression parameters for finite populations. Unpublished Ph.D. thesis, Iowa State University, Ames, Iowa.

HIDIROGLOU, M.A., FULLER, W.A., and HICKMAN, R.D. (1980). *SUPER CARP*, Dept. of Statistics, Iowa State University, Ames, Iowa.

KISH, L. (1965). *Survey Sampling*. New York: John Wiley.

LONGLEY, J.W. (1967). An appraisal of least squares programs for the electronic computer from the point of view of the user. *Journal of the American Statistical Association*, 62, 819-841.

PARK, H.J. (1987). Univariate Analysis in PC CARP. Unpublished Creative Component for the M.S., Iowa State University, Ames, Iowa.

RAO, J.N.K., and SCOTT, A.J. (1984). On chi-squared tests for multiway contingency tables with cell proportions estimated from survey data. *Annals of Statistics* 12, 46-60.

WOLTER, K.M. (1985). *Introduction to Variance Estimation*. New York: Springer-Verlag.