1980 CENSUS VARIANCE ESTIMATION PROCEDURE

Milton C. Fan, Henry F. Woltman, Susan M. Miskura and John H. Thompson, Bureau of the Census

I. INTRODUCTION

An enormous amount of data based on the 1980 census sample will ultimately be made available to the public in the form of published volumes, microfiche and computer tapes. Thus, a relatively simple yet reliable and accurate method of variance estimation is required to provide the data users with at least a measure of "sampling" error.

This paper describes the results of an empirical study designed to compare the reliability and accuracy of four commonly recommended procedures for estimating the variance of complex nonlinear estimators. Namely, the random groups, jackknife, balanced repeated replication procedures and linearization of the estimator using a Taylor series approximation in conjunction with the variance estimation formula appropriate to the specific sample design.

Each of these four basic variance estimation procedures were implemented and analyzed in this empirical study. All of these procedures were designed to provide variance estimates for 1980 census sample tabulations that will result from the raking ratio estimation procedure selected for use in the 1980 census based on the empirical study results discussed in detail in Kim, et al (1981).

As discussed in Thompson, et al (1981), three 1970 census pseudo-states comprise the empirical study populations on which the results presented here are based. These were pseudo-state 97, California counties in alphabetical order from Madera to San Diego; pseudo-state 98, California counties in alphabetical order from San Francisco to Yuba; and pseudo-state 75, Texas counties in alphabetical order from Erath to Loving. The study populations were structured into all possible systematic samples that would have resulted under the 1980 census sample design. The sampling rules for the sample design were applied to each county. Specifically, counties with less than 5,000 persons were to be sampled at the rate of 1-in-2. In the remaining counties, places of less than 5,000 persons were also sampled at a 1-in-2 rate. The remaining portions of the study population were sampled at the 1-in-6 rate. The study populations were also divided into geographically contiguous weighting areas. The variance estimation procedures were independently performed for each weighting area (WA) or in so-called pseudo-weighting areas (PWA) comprised of one or more weighting areas and for each sample.

II. VARIANCE ESTIMATORS

The four basic variance estimation procedures evaluated in this empirical study are described briefly below. A more detailed discussion of these methods is given in Thompson, et al (1981).

A. Random Groups Estimator

The random groups method of variance estimation was one of the first techniques developed to simplify variance estimation for complex survey data (Raj (1968), Kish (1965), Hansen et al (1953) and Wolter (1977)). For this procedure, the sample for a given weighting area or pseudo-weighting area is divided into g systematic mutually exclusive and exhaustive subsamples (random groups). Weighted totals are computed for each subsample (random group). Strict adherence to random groups principles, would require that the weights used to produce the weighted totals be assigned as the result of independent reweighting of each subsample using the raking ratio estimation procedure with marginal controls equal to $\frac{1}{g}$th of the total marginals. In this study both reweighting and use of the full sample weights (both derived from the raking ratio estimation procedure) were tested when the number of subsample were equal to two (i.e., when g = 2). No reweighting for values of g greater than two were done due to the budgetary constraints placed on this study. The variance for a given data item and an arbitrary weighting area is estimated using the following expression:

$$\hat{V}_{RG} = (1-f)\frac{g}{(g-1)}\sum_{j=1}^{g}\left[\hat{X}_j - \sum_{j=1}^{g}\frac{\hat{X}_j}{g}\right]^2$$

where f denotes the sampling fraction for the weighting area; $\hat{X}_j$ is the weighted total for the j^{th} subsample for the weighting area; g denotes the number of subsamples. For this study g = 2, 4, 8, 12, 16, 20 and 24.

In addition, a more conservative estimator $\hat{V}_{RGc}$, was also tested that had the following form:

$$\hat{V}_{RGc} = \hat{V}_{RG} + \frac{1}{g-1}(\hat{X} - \sum_{j=1}^{g}\hat{X}_j)^2$$

where $\hat{X}$ is the full sample total.

B. Jackknife Estimator

This method of variance estimation is a variation of the "jackknife" variance estimation technique developed by Quenouille (1956). Here, the sample is divided into g systematic subsamples. Then g pseudo-subsamples are formed, each pseudo-subsample comprising, in turn, (g-1) of the subsamples previously formed. Weighted totals are produced for each pseudo-subsample using weights which are necessarily assigned as the result of independent reweighting of each pseudo-subsample since this procedure reduces to a random groups estimator when the full sample weights are used. For a given data item and an arbitrary pseudo-weighting area, the jackknife variance estimator, denoted $\hat{V}_{JK}$, can be expressed as:

$$\hat{V}_{JK} = (1-f)\frac{g}{(g-1)}\sum_{j=1}^{g}\left[\hat{X}_{(j)} - \sum_{j=1}^{g}\frac{X_{(j)}}{g}\right]^2$$

where

f denotes the sampling fraction for the pseudo-weighting area

g denotes the number of pseudo-subsamples. For this study, g = 4,8, and 12.

$\hat{X}_{(j)}$ is the weighted total for the j^{th} pseudo-subsample based on independent weighting of the pseudo-subsample via the raking ratio estimation procedure with marginal controls equal to $\frac{g-1}{g}$th of total marginals.

C. Balanced Repeated Replication Estimator

Balanced Repeated Replication (BRR) is a method of variance estimation suggested for use in sample designs where two primaries are selected per stratum. Its use in variance estimation for such designs also employing complex nonlinear estimators has also been suggested and explored by Kish and Frankel, (1970) and McCarthy, (1966).

In this study, pseudo-weighting areas (PWAs) consisting of two weighting areas each were formed. Each sample for a given PWA is divided into a specific number, g, of subsamples. First, g-1 strata were formed of consecutive EDs. Then the sample in each stratum was divided into two systematic half-subsamples. Finally, a replicates containing exactly one of the half-subsamples from each of the strata are formed in full orthogonal balance, McCarthy (1966). Each subsample replicate is then reweighted using the census raking ratio estimation procedure with marginal controls equal to one-half of the total marginals.

To estimate the variance of any item total for an arbitrary PWA, four BRR estimators were evaluated. They are:

$$\hat{V}_{BRR} = \frac{(1-f)}{g} \sum_{\alpha=1}^{g} (2\hat{X}_\alpha - \hat{X})^2$$

$$\hat{V}^c_{BRR} = \frac{(1-f)}{g} \sum_{\alpha=1}^{g} (2\hat{X}^c_\alpha - \hat{X})^2$$

$$\hat{\bar{V}}_{BRR} = (\hat{V}_{BRR} + \hat{V}^c_{BRR}) / 2$$

$$\hat{V}^*_{BRR} = \frac{(1-f)}{g} \sum_{\alpha=1}^{g} (\hat{X}_\alpha - \hat{X}^c_\alpha)^2$$

where f is as above and g denotes the number of half-sample replicates used. The value of g considered were g=4,8, and 12.

$\hat{X}_\alpha$ denotes weighted total for the α^{th} half-sample replicates, and

$\hat{X}^c_\alpha$ denotes the weighted total for the complement of the α^{th} half-sample replicate.

$\hat{X}$ denotes the full sample weighted total for the pseudo-weighting area.

This procedure could only be performed by reweighting the half-sample replicates, as it also reduces to a random groups estimator when the full sample weights are used.

D. Linearization or Taylor Series Method

This method is frequently used to estimate the variance of a complex nonlinear estimator and is based on an approximate linearization of the estimator by a Taylor series expansion of the estimator about its expected value neglecting the higher order terms (Tepping (1968), Woodruff (1971) and Wolter (1977).)

The linearized variance estimator of an item total for single stage cluster sampling (assuming simple random sampling) and a two iteration raking ratio estimation procedure is given by Arora and Brackstone (1977). It was adapted to the proposed 1980 census raking ratio estimation procedure for this study. The form of the variance estimator is as follows:

$$\hat{V}_L(\hat{Y}) = \frac{A(A-a)}{a(a-1)} \sum_{i=1}^{R} \sum_{h=1}^{L_i} \left\{ (X_{hi.} - \frac{X_{.i.}}{n_{i.}} M_{hi.}) - \left[\sum_{j=1}^{C} \frac{\hat{X}^{(1)}_{.j}}{\hat{N}^{(1)}_{.j}} (M_{h.j} - \frac{n_{ij}}{n_{i.}} M_{hi.}) \right] \right\}^2$$

where

A is the number of housing units and GQ persons in the study population for the weighting area.

a is the number of housing units or GQ persons in the sample under considerations.

M_{hij} is the number of persons in the h^{th} sample unit who fall in the i^{th} row and j^{th} column of the weighting array. A sample unit is a housing unit or GQ person.

X_{hij} is the total number of persons in the h^{th} sample unit, i^{th} row and j^{th} column of the weighting array who possess the data item of interest.

n_{ij} is the sample person count for the cell in i^{th} row and j^{th} column of the weighting array.

N_{ij} denotes the study population person count for the cell in the i^{th} row and j^{th} column of the weighting array.

L_i denotes the number of sample units in the i^{th} row of the weighting array.

R denotes the number of rows in the weighting array.

C denotes the number of columns in the weighting array.

$$\hat{X}^{(1)}_{.j} = \sum_{i=1}^{R} X_{.ij} \frac{N_{i.}}{n_{i.}}$$

$$\hat{N}^{(1)}_{.j} = \sum_{i=1}^{R} n_{ij} \frac{N_{i.}}{n_{i.}}$$

III. ANALYSIS OF TEST RESULTS

A. Data Items and Test Runs

Fifty-seven population data items were considered covering the range of subject matter to be published in the 1980 census. The data items included seven poverty items (families and persons), eight person income items (unrelated individuals), four family income items, eight labor force items, seven educational attainment items, ten occupation items, four industry items, eight school enrollment by work status items, and one item on the number of ever married women ageds 35-44.

The portion of the study described in this paper was based on two test runs. All of the variance estimation procedures described in section II were compared in test run 1. The random groups and linearization procedures were then compared in test run 2 for PWAs consisting of only one WA. Figure 1, below lists the two test runs, the study populations considered, the number of WAs used to form PWAs, and the variance estimation procedures compared. The first two digits of the study population number indicate the 1970 census pseudo-state code. The last digit denotes the sampling rate used for the pseudo-state; a one indicates a 1-in-2 sampling rate and a two indicates a 1-in-6 sampling rate.

Figure 1 -- Summary of Test Runs Analyzed

Test Run	Study Populations	Number of Weighting Areas	Number of Weighting Areas Combined to Form Pseudo-Weighting Areas	Variance Estimation Procedures Compared
1	75-2 97-1 97-2 98-2	19 124 51 46	2	Random Groups, Jackknife, Balanced Repeated Replication, Linearization
2	75-2 97-1 97-2 98-2	19 124 51 46	1	Random Groups, Linearization

B. Methods of Comparison

The empirical or "true" (in the context of the study population) variance (VAR), and mean square error (MSE) of each variance estimation method considered were compared for every data item included in this study.

Two methods considered for comparing the statistics of interest were nonparametric methods, the Thompson-Willke test and Duncan's type multiple range test, and quantitative methods. The two methods of comparison were performed independently for MSE and VAR and for every data item included in this study. A brief description of the nonparametric tests follows.

1. Nonparametric Methods

Because the exact distribution of MSE and VAR that result from the variance estimation methods was unknown, it was felt that some type of nonparametric tests might be appropriate. The two nonparametric methods employed in this study were the Thompson-Willke Test and a Duncan's type nonparametric multiple comparison test.

a. Thompson-Willke Test

Following Youden's proposal, Thompson and Willke (1963) developed a nonparametric extreme rank sum test. This procedure was employed in this study as follows. The variance estimation methods at first were ranked by the magnitude of the measure (MSE or VAR) within each WA/PWA. Then rank sums of the methods were obtained by summing the ranks for a given method respectively over all WAs/PWAs. The test was devised to screen any method whose rank sum (or mean rank) is either large enough or small enough to be unlikely under the null hypothesis of no difference among methods. It is essentially a two-side procedure in that if the rank sum of a method for a given data item is too low at a given significance level in comparison to the expected rank sum, the null hypothesis is rejected. Independence between each WA/PWA was assumed. The validation of this assumption is unknown.

b. Duncan's Type Nonparametric Multiple Comparison Test

This Thompson-Willke Test is essentially a multiple comparisons procedure in directing the presence of extremely high or low rank sums which may result for a particular variance estimation method. In our study, it became apparent that two methods had relatively low ranks. This lead to the necessity of devising a multiple comparisons procedure to compare these two methods. This nonparametric multiple comparison test based on the Friedman rank sum was proposed and developed by Nemenyi (1963) and was discussed by Miller (1966) and Hollander and Wolfe (1973). McDonald and Thompson (1967) developed a Wilcoxon method of multiple comparison based on the range of rank sums. The Duncan's type nonparametric multiple comparison test was developed for the analysis of the results of this study. The derivation of this test is given in Fan, et al (1981). The basic credo of the test is that the difference between any two methods in a set of k methods is significantly provided the range of the rank sums of each and every subset of size p which contains the given methods is significant according to an α_p level studentized range test where p is the number of methods in the subset concerned. The test is described briefly below.

Let X_{ij} be the measure (MSE or VAR) on the i^{th} method by j^{th} WA or PWA, $i=1,\dots,k$, $j=1,\dots,n$.

The methods are ranked within each WA/PWA. Let R_{ij} be the rank of X_{ij} relative to the ordered measures $X_{(1)j} < X_{(2)j} < \dots < X_{(k)j}$ in j^{th} WA/PWA. Then mean ranks $\bar{R}_{i.}$, $i=1,2,\dots,k$ are computed.

The α_p level studentized range test is conducted by comparing the range of mean ranks of p methods involved with the critical value $q(\alpha_p,p,\infty)\left[\frac{k(k+1)}{12n}\right]^{\frac{1}{2}}$, where $q(\alpha_p,p,\infty)$ is the upper α_p point of the studentized range distribution with (p,∞) for parameters. Nemenyi's nonparametric multiple comparison test has critical value $q(\alpha,k,\infty)\left[\frac{k(k+1)}{12n}\right]^{\frac{1}{2}}$. It is essentially a Tukey type nonparametric multiple comparison test. Carmer and Swanson (1973) showed that Tukey's test is less appropriate than Duncan's multiple range test based on the results of a computer simulation study on type I error rates (i.e., probability of rejecting a true null hypothesis), type III error rates (i.e., probability of declaring one method superior to another when the reverse is actually true), and the correct decision rates (i.e., probability of declaring one method superior to another when it actually is). Hence Duncan's nonparametric multiple comparison may be more appropriate than Nemenyi's test.

IV. SUMMARY OF THE TEST RESULTS

Due to space limitations, the results from the nonparametric and quantitative analyses are only discussed briefly below. A more detailed discussion of these results is given in Fan, et al (1981).

A. Nonparametric Analysis

1. Summary of Thompson-Willke Test Results

Tables 1 and 2 summarize the results from the Thompson-Willke tests based on VAR for test runs 1 and 2, respectively using a significance level of 20-percent.

a. Test run 1 - Pseudo-Weighting Areas Consisting of Two Weighting Areas - All Methods

The summary data in Table 1.A indicate that the variance of the random groups procedure with 20 or 24 subsamples and the linearization procedure ranked significantly lower than that of the other methods for a high proportion of the total data items. The random groups, jackknife and balanced repeated replications with 2 or 4 subsamples/replicates generally yielded a fairly large number of data items having a significantly high mean rank for VAR.

b. Test run 2 - Individual-Weighting Areas - Linearization versus Random Groups

Again, on the basis of the summary data given in Fan, et al (1981) the linearization method has a significantly low mean rank for virtually every data item.

2. Summary of Duncan's Type Nonparametric Multiple Comparison

The results of the Thompson-Willke test indicated that the linearization method had a significantly low rank sum on VAR for almost every data item in both test runs. Thus a comparison of the linearization method versus the other methods was done using a Duncan's type nonparametric multiple range test. The results of this multiple comparison are summarized in Tables 3A for comparisons based on VAR for test runs 1 and 2, respectively.

a. Test run 1 - Pseudo-Weighting Areas Consisting of Two Weighting Areas - All Methods

As summarized in Table 3.A, the jackknife and balanced repeated replication methods had very few data items for which the difference in mean ranks as compared with the linearization method was not significant. For the random groups procedure, the number of nonsignificant differences increased as the number of subsamples used increased but only reached a maximum of 53 data items (out of 171) when 24 subsamples were used.

b. Test run 2 - Individual Weighting Areas - Linearization versus Random Groups

When the comparison was made between the random groups procedure and linearization method at the weighting area level, the results were fairly consistent with those from test run 1 for VAR.

In summary, the nonparametric analysis indicated that the linearization method tended to be superior to all other methods in term of variance except possibly for the random groups method using 24 subsamples. Our next objective was to obtain some quantitative measure of the magnitude of the differences (for these two methods in particular).

B. Quantitative Analysis

The quantitative analysis was performed by comparing the mean square errors that resulted from the random groups method with 24 subsamples and the linearization method. The analysis was done only for test run 2 at the one weighting area per PWA level. The analysis was carried out independently by the pseudo-states included in the test run. One comparison examined the ratio of the MSE of the random groups procedures (24 subsamples) to the MSE of the linearization method.

Table 5 -- Average Median of Ratio of Root MSEs for Selected Data Item Groups by Pseudo-State for Random Groups with 24 Subsamples to Linearization

Sampling Rate/Pseudo-State Data Item Group	1-in-2 Sampling Rate	1-in-6 Sampling Rate		
	97-1 (California)	97-2 (California)	98-2 (California)	75-2 (Texas)
Income for Unrelated Individuals	1.01	1.06	1.06	1.04
Poverty-Families and Persons	1.08	1.17	1.18	1.80
Industry and Occupation	1.08	1.20	1.22	1.53
Labor Force Status	1.08	1.19	1.22	1.29
Persons 16-21: School Enrollment by Work Status	0.93	1.03	1.01	1.00
Education for Persons 25+	1.19	1.67	1.24	1.83
Women 35-44 Ever Married	4.71	9.15	8.05	6.62
Additional Occupation	1.10	1.21	1.21	1.21
Persons 16-21 by Work Status	1.05	1.12	1.12	1.07
Family Income	1.15	1.30	1.25	1.35

The square root of the ratio of the random group MSE to the linearization MSE was computed for each weighting area. The figures in Table 5 represent the median such ratio over all weighting areas in a pseudo-state. As may be seen in Table 5, this analysis showed the linearization method to be somewhat superior to the random groups procedures, for both sampling rate categories and for most data item groups. The extreme ratios for the data of Women 35-44 ever married is due essentially to a very large positive bias that resulted for the random groups methods. The cause of this bias is still under investigation.

Further investigations into the components of the MSE - the variance and bias produced a rather unusual finding. Namely, the variance of the linearization variance estimator does not increase as the size of the variance to be estimated increases. This finding was very surprising since the variance of a total almost invariably increases as the total increases. Fan, et al (1981) illustrate this point for various data item groups for pseudo-states 75-2 and 97-2. Hence, this observation raises some questions as to the validity of using the empirical study findings to predict the variance properties of the linearization method when applied to the 1980 census sample data.

A second finding of interest concerns the bias of the linearization and random group procedures. As shown in Fan, et al (1981), the absolute bias for both the linearization and random groups procedures are roughly the same, and are relatively large in many cases. Thus, neither estimator appears to have any superior bias-reducing properties.

C. Approximate Cost of Variance Estimation Procedures

The associated costs of performing the linearization method; the random groups procedure, based on 24 subsamples; and a version of the jackknife and balanced repeated replications procedures are given in Fan, et al (1981). Currently, it is planned to produce variance estimates for about 1,000 data items tabulated from 1980 census sample data. Using the cost figures and record counts given in Fan, et al (1981), it is estimated that the cost will be approximately $750,000 to $1,000,000 to produce these estimates using the linearization method. Equivalent costs for the random groups method range from $40,000 to $90,000. Thus, the linearization method will be substantially more expensive to use than the random groups method.

V. CONCLUSIONS

In conclusion, the decision to use the random groups procedure to produce the census variance estimates was based on both cost and reliability considerations. The findings indicated that the linearization method was superior to the random groups methods, but at a substantially higher cost. Subsequent quantitative analysis indicated that the superiority of the linearization method was due almost entirely to its smaller variance. The strange findings about the distribution of the variance have cast some doubt as to whether similar results may be expected if the method were applied to the 1980 census sample data. Furthermore, both methods showed a tendency to exhibit relatively large biases, an undesirable property for any method. In our view, these findings did not by any means justify the substantial added expense of using the linearization method for what would probably have been little gain.

It should be noted that for cost purposes, consideration was given to using the linearization estimator, but only for a small subsample of the census sample. This idea was rejected, since the extent of this subsampling would without doubt make the resulting variance estimates less precise than those that could be produced from the random groups method with 24 subsamples using the entire sample.

* *

REFERENCES

Arora, H.R. and Brackstone, G.J. An Investigation of the Properties of Raking Ratio Estimators with Cluster Sampling. Technical Report, Statistics Canada, 1977

Carmer, S.G. and Swanson, M.R. An Evaluation of Ten Pairwise Multiple Comparison Procedures by Monte Carlo Methods, Journal of American Statistical Association, Vol. 68, pp. 66-74, 1973

Duncan, D.B. Multiple Range Tests for Correlated and Heteroscedastic Means, Biometrics, Vol. 13, pp. 164-176, 1957

Fan, M.C., Woltman, H.F., Miskura, S.M., and Thompson, J.H. 1980 Census Variance Estimation Procedure, presented at American Statistical Association Annual Meeting, 1981.

Hansen, M.H., Hurwitz, W.N. and Madow, W.G. Sample Survey Methods and Theory. John Wiley and Sons, Inc., 1953

Hollander, M. and Wolfe, D.A. Nonparametric Statistical Methods, John Wiley and Sons, Inc., 1973

Kim, J., Woltman, H.F., and Thompson, J.H. Analysis of the 1980 Census Sample Estimation Study Empirical Results, to appear in 1981 proceedings of section on Survey Research Methods-American Statistical Association

Kish, L. Survey Sampling. John Wiley and Sons, Inc., 1965

Kish, L. and Frankel, M.R. Balanced Repeated Replications for Standard Errors, Journal of American Statistical Association, Vol. 65, pp. 1071-1094, 1970

McCarthy, P.J. Replications: An Approach to the Analysis Data from Complex Surveys; Washington, U.S. Department of Health, Education and Welfare, Public Health Service, Publication No. 108, Series 2, Number 14, 1966

McDonald, B.J. and Thompson, W.A. Rank Sum Multiple Comparisons in One-and two-way Classification, Biometrika, Vol. 54, pp. 487-497, 1967

Miller, R.G. Simultaneous Statistical Inference, McGraw-Hill Book Company, 1966

Nemenyi, P. Distribution-free Multiple Comparisons, Unpublished Doctoral Thesis, Princeton University, Princeton, N.J., 1963

Quenouille, M.H. Notes on Bias in Estimation, Biometrika, Vol. 43, pp. 353-60, 1956

Raj, D. Sampling Theory, McGraw-Hill, Inc., 1968

Tepping, B.J. The Estimation of Variance in Complex Surveys. Proceedings of the Social Statistics Section of the American Statistical Association, pp. 11-18, 1968

Thompson, J.H., Woltman, H.F., Miskura, S.M., and Bounpane, P.A. 1980 Census Estimation and Variance Estimation Studies-Design and Methodology, presented at American Statistical Association Annual Meeting, 1981.

Thompson, W.A., Jr. and Willke, T.A. On An Extreme Rank Sum Test for Outliers, Biometrika, Vol. 50, pp. 375-383, 1963

Wolter, K.M. Introduction to Variance Estimation, Class notes for a short course in Variance Estimation given at U.S. Bureau of Census, 1977

Woodruff, R.S. A Simple Method for Approximating the Variance of a Complicated Estimate, Journal of American Statistical Association, Vol. 66, pp. 411-414, 1971

ACKNOWLEDGEMENTS

The authors would like to acknowledge the assistance of several other members of the Statistical Methods Division, Bureau of the Census for their help in preparing this paper. In particular, we wish to thank Don Luery and Paul Biemer for their comments, Paul Pavnica, Makey Lewis, David Kriegman, and Robert Jewett for the necessary computer programming, Steve Vajs for his work during the implementation and preliminary analysis phase of this study, and Jay Kim for his assistance in preparing the quantitative analysis results.

Table 1.A — Summary of Thompson-Willke Tests for Extreme Mean Ranks Based on VAR - Test Run #1

No. of Sub-samples/rep-licates	Reweighting of subsamples/replicates	Poverty (21 items) No. High 1/	Poverty (21 items) No. Low 2/	Personal Income (24 items) No. High 1/	Personal Income (24 items) No. Low 2/	Family Income (12 items) No. High 1/	Family Income (12 items) No. Low 2/	Labor Force (24 items) No. High 1/	Labor Force (24 items) No. Low 2/	Educational Attainment (21 items) No. High 1/	Educational Attainment (21 items) No. Low 2/	Occupation and Industry (42 items) No. High 1/	Occupation and Industry (42 items) No. Low 2/	School Enrollment and Work Status (24 items) No. High 1/	School Enrollment and Work Status (24 items) No. Low 2/	Women 36 – 44 Ever Married (3 items) No. High 1/	Women 36 – 44 Ever Married (3 items) No. Low 2/	Total (171 items) No. High 1/	Total (171 items) No. Low 2/
a. Random Groups																			
2	Yes	16	0	15	1	11	0	5	0	18	0	32	0	7	0	0	0	104	1
2*	Yes	18	0	17	0	12	0	11	0	21	0	34	0	11	0	0	0	124	0
2	No	15	0	16	0	11	0	16	0	20	0	36	0	9	0	3	0	126	0
4	No	6	0	2	0	8	0	9	0	17	0	24	0	2	0	3	1	71	1
8	No	0	0	0	0	0	0	7	0	2	0	0	0	0	1	3	1	12	2
12	No	0	2	0	5	2	0	2	0	0	1	0	10	0	2	3	1	7	21
16	No	0	14	0	4	4	3	1	1	0	4	0	28	0	7	3	0	8	61
20	No	0	15	0	9	4	5	2	0	0	10	0	34	0	9	3	0	9	82
24	No	0	16	0	8	4	7	0	0	0	17	0	37	0	8	3	0	8	93
b. Jackknife																			
4	Yes	6	0	3	0	8	0	0	0	9	0	16	2	4	3	0	0	46	5
8	Yes	0	0	1	0	0	0	1	1	0	0	0	0	2	0	0	2	4	3
12	Yes	0	5	1	3	0	1	3	3	0	10	0	6	4	1	0	2	8	31
c. Balanced Repeated Replication (BRR)																			
4	Yes	8	0	6	0	6	0	2	0	13	0	28	0	6	0	0	0	69	0
8	Yes	0	0	0	0	0	1	1	0	0	0	0	0	1	0	0	0	2	1
12	Yes	0	0	0	0	0	1	0	1	0	6	0	2	0	0	0	2	0	12

Table 1.A — Summary of Thompson-Willke Tests for Extreme Mean Ranks Based on VAR-Test Run #1 (Cont'd)

No. of Sub-samples/rep-licates	Reweighting of subsamples/replicates	Poverty (21 items) No. High 1/	Poverty (21 items) No. Low 2/	Personal Income (24 items) No. High 1/	Personal Income (24 items) No. Low 2/	Family Income (12 items) No. High 1/	Family Income (12 items) No. Low 2/	Labor Force (24 items) No. High 1/	Labor Force (24 items) No. Low 2/	Educational Attainment (21 items) No. High 1/	Educational Attainment (21 items) No. Low 2/	Occupation and Industry (42 items) No. High 1/	Occupation and Industry (42 items) No. Low 2/	School Enrollment and Work Status (24 items) No. High 1/	School Enrollment and Work Status (24 items) No. Low 2/	Women 36 – 44 Ever Married (3 items) No. High 1/	Women 36 – 44 Ever Married (3 items) No. Low 2/	Total (171 items) No. High 1/	Total (171 items) No. Low 2/
d. BRR - Complementary																			
4	Yes	6	0	7	0	0	0	2	0	11	0	22	0	8	0	0	0	63	0
8	Yes	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1	1
12	Yes	0	0	0	1	0	0	0	1	0	1	0	1	0	0	0	2	0	6
e. BRR - Average																			
4	Yes	5	0	4	0	8	0	2	0	10	0	21	0	5	0	0	0	55	0
8	Yes	3	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0
12	Yes	0	0	0	1	0	3	0	3	0	4	0	6	0	0	0	3	0	20
f. BRR - Difference																			
4	Yes	3	0	2	0	9	0	0	1	5	0	18	0	3	0	0	0	40	1
8	Yes	0	0	0	0	0	0	0	7	1	0	0	0	0	1	0	3	1	11
12	Yes	0	2	0	5	0	5	0	10	0	5	0	7	0	1	0	3	0	38
g. Linearization																			
NA	No	0	21	0	21	0	12	0	24	0	14	0	42	0	20	0	3	0	157

1/ Number of significantly high mean ranks at α = 0.20
2/ Number of significantly low mean ranks at α = 0.20
* Includes squared difference of average subsample total and full sample total.

Table 3A -- Summary of Duncan's Type of Multiple Comparison Test For Linearization Method Versus All Other Methods Based on VAR - Test Run #1

Number of Data Items Not Significantly Different from Linearization Method at α = 0.05

No. of Sub-samples/rep-licates	Reweighting of subsamples/replicates	Poverty (21 items)	Personal Income (24 items)	Family Income (12 items)	Labor Force (24 items)	Educational Attainment (21 items)	Occupation and Industry (42 items)	School Enrollment and Work Status (24 items)	Women 36 – 44 Ever Married (3 items)	Total (171 items)
a. Random Groups										
2	Yes	0	3	0	2	0	0	10	0	15
2*	Yes	0	3	0	1	0	0	5	0	9
2	No	0	3	0	2	0	0	8	0	13
4	No	0	6	0	2	0	0	11	0	19
8	No	1	6	0	0	0	0	10	0	17
12	No	0	10	0	1	0	1	11	0	23
16	No	2	10	0	5	0	6	11	0	34
20	No	4	14	0	2	0	9	13	0	42
24	No	7	15	1	3	2	17	10	0	53
b. Jackknife										
4	Yes	0	4	0	2	0	0	7	0	13
8	Yes	0	3	0	1	0	0	4	0	8
12	Yes	0	4	0	1	0	0	4	0	9
c. Balanced Repeated Replication (BRR)										
4	Yes	0	2	0	1	0	0	5	0	8
8	Yes	0	3	0	1	0	1	4	0	9
12	Yes	0	2	0	1	0	0	7	0	10

Table 3A -- Summary of Duncan's Type of Multiple Comparison Test for Linearization Method Versus All Other Methods Based on VAR - Test Run #1 (Cont'd)

Number of Data Items Not Significantly Different from Linearization Method at α = 0.05

No. of Sub-samples/rep-licates	Reweighting of subsamples/replicates	Poverty (21 items)	Personal Income (24 items)	Family Income (12 items	Labor Force (24 items)	Educational Attainment (21 items)	Occupation and Industry (42 items)	School Enrollment and Work Status (24 items)	Women 36 – 44 Ever Married (3 items)	Total (171 items)
d. BRR - Complementary										
4	Yes	0	3	0	2	0	0	5	0	10
8	Yes	0	4	0	1	0	1	5	0	11
12	Yes	0	2	0	0	0	3	7	0	12
e. BRR - Average										
4	Yes	0	3	0	1	0	0	7	0	11
8	Yes	0	3	0	1	0	1	4	0	9
12	Yes	0	2	0	0	0	3	6	0	11
f. BRR - Difference										
4	Yes	0	3	0	2	0	0	10	0	15
8	Yes	0	6	0	5	0	1	11	0	23
12	Yes	0	6	0	5	0	4	11	2	28

* Includes squared difference of average subsample total and full sample total.