Improved Variance Estimators Using Weighting Class Adjustments for Sample Survey Nonresponse

Shelton Maurice Jones

Inst. of Statistics Mimeo Series # 1607

ABSTRACT

JONES, SHELTON MAURICE. Improved Variance Estimators Using Weighting Class Adjustments for Sample Survey Nonresponse. (Under the direction of Robert G. D. Steel.)

In large-scale sample surveys, there will always be the problem of nonresponse. There are several imputation procedures designed to adjust for nonresponding sampling units. Among these are weighting class adjustments. One of the methods under discussion deals with weighting up responding information in class c to the total sample for the class by t_{1c}/t_{2c} where:

i) t_{1c} is an estimator of the total number of eligibles in weighting class c, and

ii) t_{2c} is an estimator of the total number of responding eligibles in weighting class c.

When this is done the estimator is no longer unbiased and common variance formulas cannot be utilized efficiently. Emphasis is placed on estimating domain totals with two estimators. The first-order Taylor series approximation or linearization method is used to obtain variance estimates under the following conditions:

Condition 1: When t_{1c} and t_{2c} are generated from sample information but are treated as constants.

Condition 2: When t_{1c} and t_{2c} are generated from sample information and both treated as estimators.

Condition 3: When t_{1c} is assumed known from an external source and treated as a constant, whereas t_{2c} is treated as an estimator based on sample information.

Condition 4: When t_{1c} is assumed known from an external source and t_{2c} an estimator based on sample information but both are treated as constants.

Variance estimates from many surveys have in the past relied on estimates that were based on Conditions 1 and 4. It has been shown, utilizing a simulation study, that variance estimates based on Conditions 1 and 4 are very erratic in accordance to variations among data, type of weighting class used and response rates. Conversely, when Conditions 2 and 3 are implemented the variance estimates are far more stable and reliable.

IMPROVED VARIANCE ESTIMATORS USING WEIGHTING CLASS ADJUSTMENTS FOR SAMPLE SURVEY NONRESPONSE

by

Shelton Maurice Jones

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Master of Science

DEPARTMENT OF STATISTICS

Raleigh

1 9 8 1

APPROVED BY:

Chairman of Advisory Committee

BIOGRAPHY

Shelton M. Jones was born near Enfield, North Carolina on May 13, 1952. The youngest of three children, he was reared in White Oak community near Enfield (Halifax County). He attended White Oak Elementary School (1958-1966) and graduated from Eastman High School (1970). He was an active member of the varsity basketball team, band, school patrols, and various other organizations.

He began financially preparing himself for study at an accredited university by working various jobs in North Carolina and Maryland during the summer of 1970. In August 1970, he entered North Carolina Central University, Durham, North Carolina, and received a Bachelor of Science degree with a major in Mathematics and a minor in Psychology, May 1974. He earned approximately 75% of his college fees working as a Teacher Assistant (1970-1973), North Carolina Central University; a Biological Aide (October 1972-May 1974), National Institute of Environmental Health Sciences, Research Triangle Park, NC; and a tutor of "slow learners" (1974), Durham Business College, Durham, NC.

From September 1974 to August 1976, he was employed as a Data Technician at the Duke University Medical Center, Durham, NC.

The author entered the Department of Statistics at North Carolina State University (NCSU), August 1976, and has worked as a Teaching and Research Assistant (January 1977-May 1979) while working toward the successful completion of requirements for the Master of Science in Statistics degree. During the summer of 1977 and 1978 he worked as a

Mathematical Statistician in the Biometry Branch, National Institute for the Environmental Health Sciences, Research Triangle Park, NC.

In June 1979 he was employed as a Statistician with the Sampling Research and Design Center (SRDC), Research Triangle Institute (RTI), Research Triangle Park, NC; and currently is employed by same.

The author is married to the former Katie Evans and is the proud father of two daughters. Mrs. Jones graduated from North Carolina Central University in May 1975 and holds a Bachelor of Science degree in Home Economics (Clothing and Textiles).

ACKNOWLEDGEMENTS

The author wishes to acknowledge with sincere appreciation the people who have advised and assisted in the preparation of this thesis.

The research problem reported in this thesis was suggested by Dr. James R. Chromy, Director of SRDC and Adjunct Professor at NCSU, and was begun under his guidance. The author therefore extends his special thanks to Dr. Chromy for his invaluable advice and guidance.

To Dr. Robert G. D. Steel, Advisor and Chairman of his Advisory Committee, the author extends his deepest gratitude, appreciation and very special thanks for the counsel, advice, guidance and continuous encouragement provided from the very beginning of his graduate study.

To Dr. L. O. Chung, Department of Mathematics and Advisory Committee member, the author extends his sincere appreciation.

The author also recognizes with many thanks Dr. Daniel G. Horvitz, Vice President of Group I, Statistical Sciences (RTI) and Mr. Stephen R. Williams, Senior Statistician, SRDC, RTI, who allowed him to reduce work hours and commitments in order to prepare this thesis.

Special thanks are extended to Dr. John C. Koop, Senior Scientist with RTI, for personal communications and verbal support given during the preparation of this thesis. Also, to other SRDC employees who provided literature in support of this work.

The author recognizes with sincere appreciation the Institute for Research in the Social Sciences, University of North Carolina at Chapel Hill for granting use of their data base which contains the County and

City Data files. Also, the Engineering Foundation for initial financial support via an entrance fellowship.

To his parents, Mr. & Mrs. Edward M. Jones, and his in-laws, Mr. & Mrs. Fred Evans, thanks for your steadfast love, encouragement and understanding. Special thanks are extended to his brother, Roger, who advised and encouraged him to undertake mathematics at the undergraduate level which led to an interest in statistics. Also, thanks for the "lively" statistical discussions.

Special thanks to Mrs. M. H. Tilley for her most valuable assistance in baby-sitting.

Finally, the author extends his warmest appreciation and love to his wife, Katie, and his daughters, Alyssa and Constance, for their understanding, patience, emotional support, encouragement, and sacrifices, without which I could never have undertaken or completed this study. Also, to his wife, Katie, thanks for typing this thesis.

TABLE OF CONTENTS

Page

LIST OF TABLES vii

1. INTRODUCTION 1

2. LITERATURE REVIEW 4

2.1 Weighting Class Adjustments 4
2.1.1 General Approach with Simple Random Sampling 4
2.1.2 Estimators Based on External Information (Sometimes Called Post-Stratification). . . 7
2.1.3 Estimators Based on Sample Information of All Eligibles in a Weighting Class 8
2.2 Estimating Domain Totals 9
2.3 Taylor Series Variance Estimation 10
2.4 Sampling Units Selected PPS With Replacement 11

3. THEORETICAL FORMULATION 14

4. EMPIRICAL STUDY DESIGN 19

4.1 Weighting Class Adjustments Used 20

5. SAMPLING FRAME DESCRIPTION 22

5.1 Response-Nonresponse Indicators 22
5.2 Domains 24
5.3 Imputations for Missing Data 24

6. RESULTS FROM SAMPLE INFORMATION 28

6.1 Comparison of Estimates for Totals 28
6.1.1 Estimates for Totals Using Sum of Weights Adjustment 29
6.1.2 Estimates for Totals Using Population Weighting Class Adjustment 29
6.2 Individual Sample Variance Comparison Using Sum of Weights Adjustment. 29
6.3 Individual Sample Variance Comparison Using Population Weighting Class Adjustment 40
6.4 Average Variance Comparison Over 100 Samples 42
6.4.1 Using Sum of Weights Adjustment 42
6.4.2 Using Population Weighting Class Adjustment 46

7. RESULTS FROM EXTERNAL INFORMATION 55

8. SUMMARY 66

LIST OF REFERENCES 68

LIST OF TABLES

Page

6.1 Comparison of Estimated Totals Over 100 Samples to the "True" Total Generated from Sample Information Using Sum of Weights Adjustment (Response Rates: 50% & 70%) . . 30

6.2 Comparison of Estimated Totals Over 100 Samples to the "True" Total Generated from Sample Information Using Sum of Weights Adjustment (Response Rates: 70% & 90%) . . . 31

6.3 Comparison of Estimated Totals Over 100 Samples to the "True" Total Generated from Sample Information Using Sum of Weights Adjustment (Response Rates: 90% & 95%) . . . 32

6.4 Comparison of Estimated Totals Over 100 Samples to the "True" Total Generated from Sample Information Using Sum of Weights Adjustment (Response Rates: 100% & 100%) . . 33

6.5 Comparison of Estimated Totals Over 100 Samples to the "True" Total Generated from Sample Information Using Population Weighting Class Adjustment (Response Rates: 50% & 70%) 34

6.6 Comparison of Estimated Totals Over 100 Samples to the "True" Total Generated from Sample Information Using Population Weighting Class Adjustment (Response Rates: 70% & 90%). 35

6.7 Comparison of Estimated Totals Over 100 Samples to the "True" Total Generated from Sample Information Using Population Weighting Class Adjustment (Response Rates: 90% & 95%) 36

6.8 Comparison of Estimated Totals Over 100 Samples to the "True" Total Generated from Sample Information Using Population Weighting Class Adjustment (Response Rates: 100% & 100%) 37

6.9 Distribution-Free Sign Test Generated from Sample Information Using Sum of Weights Adjustment. 39

6.10 Distribution-Free Sign Test Generated from Sample Information Using Population Weighting Class Adjustment. 41

6.11 Comparison of Variance Estimates Over 100 Samples Generated from Sample Information Using Sum of Weights Adjustment (Response Rates: 50% & 70%) 43

Page

6.12 Comparison of Variance Estimates Over 100 Samples Generated from Sample Information Using Sum of Weights Adjustment (Response Rates: 70% & 90%). 45

6.13 Comparison of Variance Estimates Over 100 Samples Generated from Sample Information Using Sum of Weights Adjustment (Response Rates: 90% & 95%) 47

6.14 Percent of Variance Estimates to Unbiased Estimates Generated from Sample Information Using Sum of Weights Adjustment. 48

6.15 Comparison of Variance Estimates Over 100 Samples Generated from Sample Information Using Sum of Weights Adjustment (Response Rates: 100% & 100%). 49

6.16 Comparison of Variance Estimates Over 100 Samples Generated from Sample Information Using Population Weighting Class Adjustments (Response Rates: 50% & 70%) 50

6.17 Comparison of Variance Estimates Over 100 Samples Generated from Sample Information Using Population Weighting Class Adjustment (Response Rates: 70% & 90%) 51

6.18 Comparison of Variance Estimates Over 100 Samples Generated from Sample Information Using Population Weighting Class Adjustment (Response Rates: 90% & 95%) 52

6.19 Comparison of Variance Estimates Over 100 Samples Generated from Sample Information Using Population Weighting Class Adjustment (Response Rates: 100% & 100%) 53

6.20 Percent of Variance Estimates to Unbiased Estimates Generated from Sample Information Using Population Weighting Class Adjustment. 54

7.1 Comparison of the Estimates to the "True" Total Over 100 Samples Based on External Information Using Sum of Weights Adjustment (Response Rates: 50% & 70%) 56

7.2 Comparison of the Estimates to the "True" Total Over 100 Samples Based on External Information Using Sum of Weights Adjustment (Response Rates: 70% & 90%) 57

Page

7.3 Comparison of the Estimates to the "True" Total Over 100 Samples Based on External Information Using Sum of Weights Adjustment (Response Rates: 90% & 95%) 58

7.4 Comparison of the Estimates to the "True" Total Over 100 Samples Based on External Information Using Sum of Weights Adjustment (Response Rates: 100% & 100%) 59

7.5 Comparison of the Estimated Totals to the "True" Total Based on External Information Using Population Weighting Class Adjustment (Response Rates: 50% & 70%). 60

7.6 Comparison of the Estimated Totals to the "True" Total Based on External Information Using Population Weighting Class Adjustment (Response Rates: 70% & 90%) 61

7.7 Comparison of the Estimated Total to the "True" Total Based on External Information Population Weighting Class Adjustment (Response Rates: 90% & 95%). 62

7.8 Comparison of the Estimated Total to the "True" Total Based on External Information Population Weighting Class Adjustment (Response Rates: 100% & 100%) 63

7.9 Distribution-Free Sign Test Based on External Information Using Sum of Weights Adjustment 64

7.10 Distribution-Free Sign Test Based on External Information Using Population Weighting Class Adjustment. 65

8.1 Conditional Summary of all Four Variance Estimators 67

1. INTRODUCTION

Kish (1965) describes nonresponse as "failure to obtain observations on some elements selected and designated for the sample; it may be due to refusals, not-at-homes, unreturned or lost questionnaires, etc." In large-scale sample surveys there will always be the problem of nonresponse.

Vaivanijkul (1961) investigated methods used in handling errors due to nonresponse where special reference was made to a survey of morbidity in Nashville, Tennessee. The findings revealed significant differences at the .01 level between respondents and nonrespondents for questions concerning:

i) condition of building in which they resided, and

ii) number of persons in household.

Significant differences at the .10 level were also discovered for questions concerning:

i) number of families in the household,

ii) number of deaths reported,

iii) amount of rent, and

iv) whether or not there was an inside toilet.

The remaining 28 questions did not prove to be significantly different. Surveys which consist of similar type questions should be accompanied by appropriate procedures to impute nonresponse.

There are several imputation procedures in survey sampling that are designed to adjust for nonresponding sampling units. Nonresponse introduces a source of bias upon the estimator and some questionable contribution is made to the variance of the estimator. According to the U.S.

Bureau of the Census, Current Population Survey (CPS) (1963) there is no known unbiased or even consistent method of making adjustments for nonresponse.

Common imputation methods divide the sample by survey characteristics or weighting classes. Chapman (1976) indicated that in some cases, the weights of respondents in each cell are "weighted up" to the "known" or estimated total for the cell. For a specified weighting class, the process involves a constant, or in some sample surveys one or two estimators. The estimators are as follows:

i) an estimator of the total number of eligibles in weighting class c, t_{1c},

ii) an estimator of the total number of responding eligibles in weighting class c, t_{2c}.

In most sample surveys t_{2c} will be an estimator. And thus, the weighting class adjustment is t_{1c}/t_{2c}. Quite often, classical procedures treat both estimators as constants. It is not clear what their real contributions are to the sampling variance.

The current paper deals with approximating the variance of two estimators for totals when using weighting class adjustments for imputing nonresponse. The first-order Taylor series approximation formula is used to estimate this variability under the following conditions:

Condition 1: When t_{1c} and t_{2c} are generated from sample information but are treated as constants.

Condition 2: When t_{1c} and t_{2c} are generated from sample information and both treated as estimators.

Condition 3: When t_{1c} is assumed known from an external source and treated as a constant, whereas t_{2c} is treated as an estimator based on sample information.

Condition 4: When t_{1c} is assumed known from an external source and t_{2c} an estimator based on sample information but both are treated as constants.

Each of the above conditions will yield an estimate of the true variance. A simulation study is utilized to compare these estimates with the variance estimate of the statistics based on repeated samples.

Many sample surveys rely heavily upon variance estimates that are based on conditions 1 and 4. There has been no mention of conditions 2 and 3 in the literature search. It shall be shown that conditions 2 and 3 yield estimates that have better properties and are far more reliable than 1 and 4.

2. LITERATURE REVIEW

2.1 Weighting Class Adjustments

2.1.1 General Approach with Simple Random Sampling

A weighting class adjustment for nonresponse may be defined as a weighting up of the responses within class c to the total sample for the class. It is assumed that respondents have provided information which is representative of the nonrespondents. The objective for using weighting class adjustments as well as other imputation methods, is to reduce the nonresponse bias, assuming the nonrespondents respond in a similar manner.

Now to examine the bias based on formulation by Chapman that is associated with a simple random sample without replacement. The sample weight for the c^{th} weighting class is N/n and the nonresponse adjustment is n_c/n_{cr},

where N = the total number of sampling units in the population

n = the total number of sampling units in the sample

n_c = the total number of sampling units from the c^{th} weighting class in the sample,

n_{cr} = the total number of responding sampling units from the c^{th} weighting class in the sample.

According to Chapman, an estimate of the population total for the variable x is:

$$\hat{F}_r = N \sum_{c=1}^{C} p_c \bar{x}_{cr} \qquad (2.1)$$

where: C = the total number of weighting classes,

$\bar{x}_{cr}$ = the sample mean among respondents in the c^{th} weighting class, and

p_c = the proportion of the sample in the c^{th} weighting class, i.e., n_c/n.

The expected value of $\hat{F}_r$ is:

$$E(\hat{F}_r) = N\ E \sum_{c=1}^{C} p_c \bar{x}_{cr}$$

$$= N \sum_{c=1}^{C} P_c \bar{X}_{cr} \qquad (2.2)$$

where: $P_c = E(p_c)$

$\bar{X}_{cr}$ = the population mean among respondents in the c^{th} weighting class.

The bias of equation 2.1 may be written as:

$$\text{bias}\ (\hat{F}_r) = E(\hat{F}_r) - F_r$$

$$= N \left(\sum_{c=1}^{C} P_c \bar{X}_{cr} - \sum_{c=1}^{C} P_c \bar{X}_c\right)$$

$$= N \sum_{c=1}^{C} P_c (\bar{X}_{cr} - \bar{X}_c) \qquad (2.3)$$

where: $\bar{X}_c$ = the population mean of all sampling units within the c^{th} weighting class.

Since $\bar{X}_c$ is composed of respondents and nonrespondents,

$$\bar{X}_c = P_{cr}\bar{X}_{cr} + Q_{cn}\bar{X}_{cn} \qquad (2.4)$$

where: $\bar{X}_{cn}$ = the population mean of nonrespondents in the c^{th} weighting class.

P_{cr} and Q_{cn} are population proportions in class c which are respondents and nonrespondents, respectively, that is:

$$P_{cr} = n_{cr}/n_c$$

$$Q_{cn} = n_{cn}/n_c \ .$$

(n_{cn} equals the total number of nonresponding sampling units from the c^{th} weight class in the sample.)

Hence, the bias of the estimator is:

$$\begin{aligned} \text{bias } (\hat{F}_r) &= N \sum_{c=1}^{C} P_c(\bar{X}_{cr} - \bar{X}_c) \\ &= N \sum_{c=1}^{C} P_c(\bar{X}_{cr} - P_{cr}\bar{X}_{cr} - Q_{cn}\bar{X}_{cn}) \\ &= N \sum_{c=1}^{C} P_c\{(1 - P_{cr})\bar{X}_{cr} - Q_{cn}\bar{X}_{cn}\} \\ &= N \sum_{c=1}^{C} P_c Q_{cn}(\bar{X}_{cr} - \bar{X}_{cn}) \ . \qquad (2.5) \end{aligned}$$

Therefore, according to Chapman, by using weighting class adjustments the nonresponse bias can be reduced if:

i) In general, $(\bar{X}_{cr} - \bar{X}_{cn}) < (\bar{X}_{.r} - \bar{X}_{.n})$, and

ii) the nonresponse proportions (Q_c's) vary over classes,

where $\bar{X}_{.r} = \sum_{c=1}^{C} P_c\bar{X}_{cr}$, and

$$\bar{X}_{.n} = \sum_{c=1}^{C} P_c\bar{X}_{cn} \ .$$

2.1.2 Estimators Based on External Information (Sometimes Called Post-Stratification)

The number of eligibles (T_{1c}) is often assumed known from the population. Knowledge may be based on a previous survey, an existing file or from some prior distribution. Whereas t_{2c} may be generated from sample results, the ratio (T_{1c}/t_{2c}) is greater than or equal to one and is the post-stratification adjustment coefficient. An example would be to look at a simple random sample (SRS) of individuals stratified by age and sex. Referring to Hansen, Hurwitz and Madow (1953), fairly accurate CPS estimates of the total population in the United States exist for an age and sex distribution. The CPS is implemented several times a year and estimates are fairly reliable. Therefore, many ongoing surveys that have post-stratified similarly, take the corresponding total for the number of eligibles and use it in their adjustment coefficients. Quite often when T_{1c} is obtained by these means it is considered a constant in variance formulation.

An estimator for the number of respondents will be expressed as:

$$t_{2c} = \sum_{k=1}^{n} \frac{N}{n} Y_{2ck} \qquad (2.6)$$

where: Y_{2ck} = 1, if sample unit k is an element of weighting class c and is a respondent,

= 0, otherwise.

In terms of variance estimation, it should be emphasized that T_{1c} and t_{2c} are generally treated as constants. In reality they are, respectively, a constant and an estimator for the number of respondents in weighting class c.

Let $\hat{F}_2$ be an estimator of totals for the variable of interest based on some previous survey or census. Then it follows that:

$$\hat{F}_2 = \sum_{c=1}^{C} \frac{T_{1c}}{t_{2c}} t_{3c} \tag{2.7}$$

where: $$t_{3c} = \sum_{k=1}^{n} \frac{N}{n} Y_{3ck} \tag{2.8}$$

Y_{3ck} = the value for the variable of interest given sample unit k is a respondent in weighting class c,

= 0, otherwise.

2.1.3 Estimators Based on Sample Information of All Eligibles in a Weighting Class

In many sample surveys, population totals for certain weighting classes are not known and must be estimated from sample results. A procedure similar to that of the U.S. Bureau of the Census consists of calculating the ratio of the total number of eligible and responding households in each weighting class from sample information. This ratio is used to "weight up" the responding household information and may be expressed as t_{1c}/t_{2c},

where: t_{1c} = sum of weights for all eligibles in weighting class c

t_{2c} = sum of weights for all responding eligibles in weighting class c

The Current Population Survey (1978) consist of dividing the units into 48 noninterview adjustment cells or classes. Cells are composed of 2 race classifications, 3 residence categories, and 8 rotation groups. In most instances, unweighted counts of households are used to adjust for noninterviewed households. The following ratio is computed for each cell in each rotation group:

$$\frac{t_{1c}}{t_{2c}} = \frac{\text{interviewed households + noninterviewed households}}{\text{interviewed households}}$$

The ratios are uniformly applied to each respondent in all cells as long as it does not exceed 2.0. If the ratio exceeds 2.0 counts are grouped for all races within the residence category in the cluster.

Classical procedures such as SESUDAAN[1] (1979) treat both t_{1c} and t_{2c} as constants in the computation of variance estimates associated with equation 2.7 and,

$$\hat{F}_1 = \sum_{c=1}^{C} \frac{t_{1c}}{t_{2c}} t_{3c} \, . \qquad (2.9)$$

2.2 Estimating Domain Totals

Domains are viewed as reporting subgroups from the population of interest. In a national education study, a domain may be the subgroup of students in public schools in grades 10 through 12. As is the case here, in domains of study the numbers of sampled units are random variables. Kendall and Stuart (1968) discussed an example of a sample from a human population where interest was in calculating statistics for two subgroups, men and women. They mention the fact that, had the sample been stratified into men and women prior to sample selection, then sample sizes for the two subgroups would have been fixed. But often it is not

[1]SESUDAAN is a software package that computes standard errors for standardized rates from sample survey data.

possible to stratify by domains before selecting the units. Kendall and Stuart footnoted that "there is a complication here in the case of non-response, since nonresponse may be correlated with the value of y (the characteristic of interest), so that the responding group cannot provide an unbiased estimator of y for the population as a whole."

According to Hartley (1959) sometimes no decision can be made as to which sampling units belong to a certain domain until after the survey has been implemented. This among others is a distinction between domains and treatment groups of experiments.

Now to continue SRS procedure described in sections 2.1.1, 2.1.2, and 2.1.3. Hence, the estimators of totals for domain j, complicated by nonresponse, may be expressed as:

$$\hat{F}_1(d) = \sum_{c=1}^{C} \frac{t_{1c}}{t_{2c}} t_{3c}(d) \quad , \quad \hat{F}_2(d) = \sum_{c=1}^{C} \frac{T_{1c}}{t_{2c}} t_{3c}(d)$$

where $t_{3c}(d) = \sum_{k=1}^{n} \frac{N}{n} Y_{3ck}(d)$ (2.10)

$Y_{3ck}(d)$ = the value of the variable of interest given sample unit k is a member of domain d and a respondent in weighting class c,

= 0, otherwise.

2.3 Taylor Series Variance Estimation

The linearization method or Taylor series variance approximation is widely used among various research agencies. Common goals are to estimate the variance associated with a nonlinear estimator by reducing it to a linear form. The form is defined by means of partial derivatives whereas second and higher terms are ignored. The National Center for Health Statistics (1975) reports that the Canadian Labor Force Survey uses linearization to obtain variance estimates of ratio-estimated characteristics. The approximation is also utilized by the U.S. Bureau of the Census in the CPS. Pertaining to statistics from the CPS, it is stated that "the variance of the linear function is a close approximation to the variance of the original expression. However, for some complicated estimators, the use of the Taylor approximation may give a poor estimate of the variance."

2.4 Sampling Units Selected PPS With Replacement

Previous emphasis has been on simple random sampling without replacement. It is more often the case that large-scale sample surveys consist of designs which are not entirely simple random but stratified having selection of first stage units with unequal probability. Further discussion will be concerned with sampling first stage units with probability proportional to size (PPS) with replacement. As emphasized by Stuart (1962) such a sampling procedure produces a self-weighting sample.

The design of interest consists of two sampling stages. It is not necessary to place any restriction on the selection scheme of secondary units as long as the procedure yields a probability sample. An advantage of this design is that there is no limit to the number of subsequent sampling stages as long as primary units are sampled PPS with replacement and other stages are independent.

The assumption is that the size measure is highly correlated with the characteristic variable of interest, hence this selection procedure should improve the efficiency of the estimate of total. This fact is mentioned in Singh (1978).

Another advantage in sampling first stage units PPS is that the form of the estimated variance is quite simple, as will be shown later. A further advantage is the variance associated with replacement may be used even when first stage units are selected without replacement. This result was substantiated by Raj (1964). He determined that the with-replacement variance was slightly greater than that of without replacement. So, if the variance associated with PPS with replacement is used when in fact first stage units were selected PPS without replacement then the resulting variance will slightly overestimate the "true" variance. Having a variance estimate which is moderately inflated creates few problems but gives the researcher some protection against making false inferences, as opposed to an underestimate.

Estimators that are to follow in this section were cited from Cochran (1977). The unbiased estimator of the population total is described as:

$$\hat{F}_{ppz} = \frac{1}{n} \sum_{k=1}^{n} \frac{\hat{F}_k}{z_k}$$

where $\hat{F}_z$ = the product of size measure k and the mean of sample unit k, i.e., $M_k \bar{f}_k$

z_k = size measure k divided by the sum of all size measures in the population, i.e., M_k/M_o

The variance of $\hat{F}_{ppz}$ is:

$$Var(\hat{F}_{ppz}) = \frac{1}{n} \sum_{k=1}^{N} z_k \left(\frac{F_k}{z_k} - F\right)^2 + \frac{1}{n} \sum_{k=1}^{N} \frac{M_k^2(1 - g_{2k})S_{2k}^2}{m_k z_k} \qquad (2.11)$$

where $F_k = E(\hat{F}_k)$

F = the population total of interest

g_{2k} = m_k sample subunits divided by M_k, i.e., m_k/M_k

S_{2k}^2 = the variance among subunits in unit k.

Therefore, according to Cochran an unbiased sample estimator of the variance of equation 2.11 is:

$$\hat{V}ar(\hat{F}_{ppz}) = \frac{\sum_{k=1}^{n} \left(\frac{\hat{F}_k}{z_k} - \hat{F}_{ppz}\right)^2}{n(n-1)} \qquad (2.12)$$

3. Theoretical Formulation

Estimators for the variance of $\hat{F}_1$ and $\hat{F}_2$ in equations 2.9 and 2.7 have complicated mathematical structures whose complexity lies in the number of estimators present and whether the covariances are zero. Clearly, the variance estimators will be in their simplest form when the number of eligibles and respondents are considered known.

An expansion of the Taylor Series will be used to reduce these nonlinear forms to linear forms. A common assumption will be made here, that is, the samples are large enough so that the remaining terms after the first-order approximations can be omitted. If the partial derivatives are evaluated at their expected values then the large sample approximation for the variance of $\hat{F}$ is:

$$\mathrm{Var}\ (\hat{F}) \doteq E \sum_{c=1}^{C} \sum_{a=1}^{3} \frac{\partial \hat{F}}{\partial t_{ac}} (T_{ac} - E\ t_{ac})^{2}$$

$$= \sum_{c=1}^{C} \sum_{a=1}^{3} \left(\frac{\partial \hat{F}}{\partial t_{ac}}\right)^2 V(t_{ac}) + \sum_{c=1}^{C} \sum_{a=1}^{3} \sum_{c'=1}^{C} \sum_{\substack{a'=1 \\ (c,\ a) \neq (c',\ a')}}^{3} \frac{\partial \hat{F}}{\partial t_{ac}} \frac{\partial \hat{F}}{\partial t_{a'c'}} \mathrm{Cov}(t_{ac} t_{a'c'}) \qquad (3.1)$$

where: $\hat{F} = \hat{F}_1$ if t_{1c} is generated based on sample information, or

$\hat{F} = \hat{F}_2$ if T_{1c} is assumed known from an external source.

In practice the partial derivatives are evaluated at the estimated values. It should be noted that if t_{ac} and $t_{a'c'}$ are independent then of course $\mathrm{Cov}(t_{ac}\ t_{a'c'}) = 0$, which implies that $\mathrm{Cov}(t_{ac}\ t_{a'c}) = 0$, and equation 3.1 reduces to:

$$\mathrm{Var}\ (\hat{F}) \doteq \sum_{c=1}^{C} \sum_{a=1}^{3} \left(\frac{\partial \hat{F}}{\partial t_{ac}}\right)^2 V(t_{ac}). \qquad (3.2)$$

Efforts will now be made towards deriving an estimator for the variance of $\hat{F}_1$ and $\hat{F}_2$ based on the conditions which were described and sample design. As revealed in section 1, the four conditions are as follows:

Condition 1: When t_{1c} and t_{2c} are generated from sample information but are treated as constants.

Condition 2: When t_{1c} and t_{2c} are generated from sample information and both treated as estimators.

Condition 3: When t_{1c} is assumed known from an external source and treated as a constant, whereas t_{2c} is treated as an estimator based on sample information.

Condition 4: When t_{1c} is assumed known from an external source and t_{2c} an estimator based on sample information but both are treated as constants.

To recall $\hat{F}_1$ and $\hat{F}_2$ in equations 2.9 and 2.7,

$$\hat{F}_1 = \sum_{c=1}^{C} \frac{t_{1c}}{t_{2c}} t_{3c} \quad , \quad \hat{F}_2 = \sum_{c=1}^{C} \frac{T_{1c}}{t_{2c}} t_{3c}$$

where for equation 2.9,

$$t_{ac} = \sum_{h=1}^{R} \sum_{k=1}^{n_h} t_{achk}, \quad a\varepsilon\{1,2,3\}, \text{ in equation 2.7, } a\varepsilon\{2,3\}$$

$$t_{achk} = \sum_{i=1}^{m_{hk}} W_{hki} Y_{achki}$$

W_{hki} = the sampling weight corresponding to SSU i within PSU k within stratum h

m_{hk} = the number of SSU's within PSU k within stratum h

Y_{1chki} = 1, if SSU i within PSU k within stratum h is in weighting class c

= 0, otherwise

Y_{2chki} = 1, if SSU i within PSU k within stratum h is in weighting class c and is a respondent

= 0, otherwise

Y_{3chki} = the value for the variable of interest given SSU i within PSU k within stratum h is in weighting class c and is a respondent

= 0, otherwise

The expansion in equation 3.1 is often tedious to work with in practice. A better procedure was demonstrated by Woodruff (1971). Thus, pertaining to Woodruff's procedure, an approximation of the variance of $\hat{F}$ may be expressed as:

$$\text{Var}\ (\hat{F}) \doteq E\left\{\sum_{c=1}^{C}\sum_{a=1}^{3}\frac{\partial\hat{F}}{\partial t_{ac}}\,(t_{ac} - E\ t_{ac})\right\}^2$$

$$= E\left\{\sum_{c=1}^{C}\sum_{a=1}^{3}\frac{\partial\hat{F}}{\partial t_{ac}}\left[\sum_{h=1}^{R}\sum_{k=1}^{n_h} t_{achk} - \sum_{h=1}^{R}\sum_{k=1}^{n_h}\frac{T_{ach}}{n_h}\right]\right\}^2 \qquad (3.3)$$

According to Woodruff if the order of summation is reversed the variance can be easily evaluated. Using this method equation 3.1 is equivalent to

$$= E\left\{\sum_{h=1}^{R}\sum_{k=1}^{n_h}\left[\sum_{c=1}^{C}\sum_{a=1}^{3}\frac{\partial\hat{F}}{\partial t_{ac}}\,t_{achk} - \sum_{c=1}^{C}\sum_{a=1}^{3}\frac{\partial\hat{F}}{\partial t_{ac}}\,\frac{T_{ach}}{n_h}\right]\right\}^2$$

$$= E\left\{\sum_{h=1}^{R}\sum_{k=1}^{n_h}\sum_{c=1}^{C}\sum_{a=1}^{3}\frac{\partial\hat{F}}{\partial t_{ac}}\,t_{achk} - \sum_{h=1}^{R}\sum_{k=1}^{n_h}\sum_{c=1}^{C}\sum_{a=1}^{3}\frac{\partial\hat{F}}{\partial t_{ac}}\,\frac{T_{ach}}{n_h}\right\}^2$$

$$= \text{Var}\left\{\sum_{h=1}^{R}\sum_{k=1}^{n_h}\sum_{c=1}^{C}\sum_{a=1}^{3}\frac{\partial\hat{F}}{\partial t_{ac}}\,t_{achk}\right\} \qquad (3.4)$$

where the partial derivatives $(\frac{\partial \hat{F}}{\partial t_{ac}})$ are evaluated at estimated values, and $T_{ach} = E\ (t_{achk})$.

At this point another variable will be defined.

$$\hat{U}_{hk} = n_h \sum_{c=1}^{C} \sum_{a=1}^{3} \frac{\partial \hat{F}}{\partial t_{ac}} t_{achk} \quad .$$

Then, equation 3.4 is equal to,

$$\text{Var} \sum_{h=1}^{R} \frac{1}{n_h} \sum_{k=1}^{n_h} \hat{U}_{hk} = \sum_{h=1}^{R} \sum_{k=1}^{n_h} \frac{(\hat{U}_{hk} - \bar{U}_h)^2}{n_h(n_h-1)} \qquad (3.5)$$

where: $\bar{U}_h$ = the average of $\hat{U}_{hk}$ over the stratum, as indicated in Cochran.

Without regard to the sample design each t_{ac} will always be an estimator for the characteristic of interest. As discussed previously to some degree, the following conditions will be explored to examine the variance of the estimators:

Condition I - t_{1c}, t_{2c} assumed constants and t_{3c} an estimator

Condition II - each t_{ac} is an estimator, where $a \in \{1,2,3\}$

Condition III - T_{1c} assumed constant and t_{2c}, t_{3c} estimators

Condition IV - T_{1c} assumed constant and each t_{ac} is an estimator, where $a \in \{2, 3\}$.

It is apparent that the variance of $\hat{F}$ as expressed in equation 3.5 is a function of $\hat{U}_{hk}$. Now to consider an expansion of $\hat{U}_{hk}$ for each condition.

Condition I

$$\hat{U}_{hk} = n_h \sum_{c=1}^{C} \frac{\partial \hat{F}_1}{\partial t_{3c}} t_{3chk} = n_h \sum_{c=1}^{C} \frac{t_{1c}}{t_{2c}} t_{3chk} \qquad (3.6)$$

Condition II

$$\hat{U}_{hk} = n_h \sum_{c=1}^{C} \sum_{a=1}^{3} \frac{\partial \hat{F}_1}{\partial t_{ac}} t_{achk}$$

$$= n_h \sum_{c=1}^{C} \frac{t_{3c}}{t_{2c}} t_{1chk} - n_h \sum_{c=1}^{C} \frac{t_{1c}t_{3c}}{(t_{2c})^2} t_{2chk} + n_h \sum_{c=1}^{C} \frac{t_{1c}}{t_{2c}} t_{3chk} \quad (3.7)$$

Condition III

$$\hat{U}_{hk} = n_h \sum_{c=1}^{C} \sum_{a=2}^{3} \frac{\partial \hat{F}_2}{\partial t_{ac}} t_{achk}$$

$$= n_h \sum_{c=1}^{C} \frac{T_{1c}}{t_{2c}} t_{3chk} - n_h \sum_{c=1}^{C} \frac{T_{1c}t_{3c}}{(t_{2c})^2} t_{2chk} \quad (3.8)$$

Condition IV

$$\hat{U}_{hk} = n_h \sum_{c=1}^{C} \frac{\partial \hat{F}_2}{\partial t_{3c}} t_{3chk}$$

$$= n_h \sum_{c=1}^{C} \frac{T_{1c}}{t_{2c}} t_{3chk} \quad (3.9)$$

Therefore, the above $\hat{U}_{hk}$'s in equations 3.6-3.9 will be utilized in equation 3.5 for the variance comparisons.

4. EMPIRICAL STUDY DESIGN

The present survey is a simulation study that is designed to obtain national estimates from a 1975 data file with adjustments made for nonresponding sampling units. A total of 100 samples are generated for the purpose of variance comparison. The design consists of 4 strata which are the 4 Census Regions. Sampling is done at 2 stages. First stage units are the 50 states and Washington, DC. Four states are selected PPS with replacement from within each stratum, with the size measure being the 1970 state population. Five counties are also selected PPS with replacement from within each state, with the size measure being the 1970 county population. Therefore, each of the 100 samples is composed of 80 second stage units from which national estimates will be obtained.

The purpose of the survey is to estimate totals for 9 variables, each having 2 domains. These totals already exist as is described in the County And City Data Book, U.S. Department of Commerce (1975). The volume contains "a variety of statistical information for counties, standard metropolitan statistical areas, cities, urbanized areas, and unincorporated places," U.S. Department of Commerce.

Below are the 9 variables of interest for the United States in 1975.

1) Births
2) Divorces
3) Population
4) Food Stores
5) Gasoline Service Stations
6) Hospitals

7) Marriages

8) Public School Enrollment

9) Motor Vehicle Thefts

4.1 Weighting Class Adjustments Used

Discussion will be confined to two different types of weighting class adjustments. The first adjustment is similar to that used by the U.S. Bureau of the Census (1963). As previously stated, the Census Bureau weights up household information within each group or class;

$$\frac{t_{1c}}{t_{2c}}, \qquad (4.1.1)$$

where: t_{1c} = interviewed households plus noninterviewed households in class c

t_{2c} = interviewed households in class c.

The weighting, of course, is done within groups or weighting classes where the ratio does not exceed 2. Likewise, in the present survey (4.1.1) is equal to:

$$\frac{\text{an estimate of the total number of counties in class c}}{\text{an estimate of the total number of responding counties in class c}}$$

$$= \frac{\text{sum of weights for all counties in class c}}{\text{sum of weights for responding counties in class c}}.$$

In the present study, this adjustment is rather general since it does not fully take into consideration the particular selection scheme used to select sampling units at both stages. States and counties were selected PPS with the 1970 population being the size measure. Thus, it is feasible that a better adjustment would be relative to the population. Also, for this design, the variance of an estimate which makes use of a population adjustment should prove to be smaller than the variance of a similar estimator which does not use the population adjustment. The population adjustment is as follows:

$$\frac{t_{1c}}{t_{2c}} = \frac{\text{an estimate of 1970 population of all counties in class c}}{\text{an estimate of 1970 population of responding counties in class c}}$$

In this study there will be two weighting classes. They will be based upon the percent of the county's population that lies in an urbanized area. Counties whose population is 50% urban or less will be included in weighting class 1 and greater than 50% urban in weighting class 2. The adjustments mentioned above will be applied within these classes.

5. SAMPLING FRAME DESCRIPTION

5.1 Response-Nonresponse Indicators

In this survey it is necessary to control the probability that any given second stage unit is a respondent. Second stage units are designated as a respondent or nonrespondent by utilizing the function "uniform," a mathematical function in Statistical Analysis Systems (SAS)(1979). Given a five-digit seed, this function is used to generate 10 random numbers for each county selected over the uniform interval (0,1). Increasing the numbers generated by a factor of ten will produce uniform random numbers (X_i) over the interval (0,10). The following probabilities will apply for each X_i:

$X_i \leq k$, (i, k) = 1,2,...9	Probability
1	.1
2	.2
3	.3
4	.4
5	.5
6	.6
7	.7
8	.8
9	.9
$X_{10} < 10$	1.0

Now, an indicator variable (Y_i) shall be created as a function of each X_i.

$$Y_i = 1;\ \text{if } X_i \leq k \quad (k=1, 2, \ldots, 9)$$
$$= 0;\ \text{otherwise}$$
$$Y_{10} = 1;\ \text{if } X_{10} < 10$$
$$= 0;\ \text{otherwise}$$

$Y_i = 1$ will imply that the county is a respondent for the i^{th} response rate,

where: $i = 1$ implies a 10% response rate

$i = 2$ implies a 20% response rate

and so on, to $i = 10$ (100% response rate). In addition $i = 95$ will imply a 95% response rate which is obtained by $Y_{95} = 1 - (Y_1 \cdot Y_5)$. So, for the i^{th} response rate the probability that a given county is a respondent may be expressed as:

$$P(Y_i=1) = P(X_i \leq k) \text{ for } (i,k) = 1,2,\ldots,9,95 \text{ and}$$
$$P(Y_{10}=1) = P(X_{10} < 10)$$

More clearly, if interest lies in achieving a 70% response rate over counties then,

$$P(Y_7=1) = P(X_7 \leq 7) = .7$$

that is, $Y_7 = 1$ approximately 70% of the time over all counties.

In a significant number of large-scale surveys response rates tend to lie between 100% and 50%. So, response rates for weighting classes will be assigned between this interval with the lower rate occurring in weighting class 1. Analyses will be confined to the following response rates:

Analysis Number	Weighting Class	Response Rate
1	1	50% Low Rate
	2	70%
2	1	70% Medium Rate
	2	90%
3	1	90% High Rate
	2	95%
4	1	100% Perfect
	2	100% Response Rate

5.2 Domains

Domain #1 consist of those counties whose 1970 population (S_{hki}), as determined by the 1970 census, was less than 135,000. Domain #2 includes counties whose 1970 population was 135,000 or greater. Out of 3143 counties in the sampling frame, 2877 are in domain #1 and the remaining 266 in domain #2.

The following is a distribution of weighting class by domain.

Weighting Class	Domains 1	Domains 2	Total
1	2168	10	2178
2	709	256	965
TOTAL	2877	266	3143

5.3 Imputations for Missing Data

This study requires all county records to have a reasonable response for each variable prior to the onset of the survey. However, there are missing data values for some counties for different variables. The only variable that does not have missing data is the population. Data is available for both 1970 and 1975.

For imputation purposes, the assumption is made that the magnitude of each variable total corresponds with the 1975[1] population of a county. For example, the total number of births for a county will increase or decrease according to the county's population. These imputations were uniformly applied within strata since the assumption is made that county characteristics within strata are more homogeneous than in the United States as a whole. Thus, a single adjustment over strata may be considered less efficient.

The totals F_v (v = 1,2,3,...9) are known for each variable on the stratum level (which count missing values as 0). Also, the stratum population (S_h) is known for each variable. A factor in the imputation procedure is then F_v/S_h. This ratio is multiplied by the county population to impute a value as follows:

$$\left(\frac{F_v}{S_h}\right) S_{hki} \quad .$$

For 1975, no data exists on the frame for the variables "food stores" and "gasoline service stations." But, there are 1972 data available for all counties and for both variables. A similar imputation method is used to impute a value for each county record. The method utilizes the 1972 data along with a "1975 adjustment factor." This factor is an adjustment that is also applied uniformly within strata and is simply the ratio of the 1975 stratum population (S_{h75}) to the 1972 stratum population (S_{h72}). Here the assumption is still

[1]In practice, it would be necessary to use the county's 1970 population as a factor towards imputing missing values since the 1975 population would be unknown and is one of the variables to be estimated.

made that these variables will increase or decrease at a comparable rate as the population. Therefore, a "frame county estimate" for either of the two variables is:

$$\left(\frac{S_{h75}}{S_{h72}}\right) X_t \quad , \quad t = 1, 2$$

where: X_1 = number of food stores in county i

X_2 = number of gasoline service stations in county i

An element of the sampling design requires four states to be selected PPS with replacement from each strata. Secondly, five counties are to be selected PPS with replacement from each state. As of 1970, both Delaware and Hawaii had fewer than five counties. Delaware has three counties and Hawaii has four. For purposes of this selection scheme, these two states will be combined with a neighboring state within the same stratum. Delaware is combined with Maryland increasing the number of counties and independent cities in Maryland from 24 to 27 and increasing the first stage selection probability from 3923892/62813102 to 4472001/62813102. Hawaii and Alaska fall in the same stratum so they will be combined increasing the number of counties in Alaska from 29 to 33 counties with first stage probability of selection increasing from 769913/34838243 to 1072496/34838243.

Washington, DC, has no counties so for similar reasons it will be combined with Virginia. This yields 137 counties and independent cities for Virginia and an increase in selection probability from 4651448/62813102 to 5408116/62813102.

Further imputations are made for counties and cities in Virginia and Maryland. There are several cases for these states where data is consolidated for the county with missing values for each independent city in the county. At any rate, to obtain a value for such missing data, a quantity is imputed which is proportional to the population of the pertinent counties or county equivalent.

6. RESULTS FROM SAMPLE INFORMATION

6.1 Comparison of Estimates for Totals

For each response rate, a paired t-test is used to determine if these estimates are really different from the "true" total. The test is a modification of that indicated in Steel and Torrie (1980). Objectives are to test:

Ho: $\mu_d = 0$

Ha: $\mu_d \neq 0$

where μ_d is the mean of the population of the differences,

$d_{1jk} = \hat{F}_{1jk} - F_{1jk}$.

$\hat{F}_{1jk} = \frac{1}{100} \sum_{i=1}^{100} \hat{F}_{1ijk}$, i.e., the average total over 100 samples.

The totals $\hat{F}_{1jk}$ and F_{1jk} are respectively for the j^{th} variable and k^{th} domain estimate, and the j^{th} variable and k^{th} domain "true" total.

From Hollander and Wolfe (1973), Fisher's distribution-free sign test will also be used to determine if the estimates of total are statistically different from the "true" total. The objectives are to test:

Ho: $\theta = 0$

Ha: $\theta \neq 0$

where: θ = median of the Z_{1jk}'s

$Z_{1jk} = \hat{F}_{1jk} - F_{1jk}$.

The mentioned procedures will be used on both types of weighting class adjustments.

6.1.1 Estimates for Totals Using Sum of Weights Adjustment

It can be seen in Table 6.1 at the low response rate that $\hat{F}_1$ tends to overestimate the "true" total. The paired t-test indicated in section 6.1 proved to be significant at the 5% level.

The Medium rate through the Perfect Response Rate (Tables 6.2, 6.3, and 6.4) yielded t values which are not statistically significant. Apparently, higher response rates tend to generate estimates that are closer to the "true" total. This conclusion is about what could be expected. Fisher's distribution-free sign test described in section 6.1 also agrees with the above conclusions.

6.1.2 Estimates for Total Using Population Weighting Class Adjustments

Using this particular type of adjustment it is obvious that at the low response rate as indicated in Table 6.5, good estimates are achieved. In fact, even at such a low response rate the estimates are extremely close to the "true" totals. The paired t-test and sign test of differences previously mentioned in section 6.1, substantiate this result and thus determines the estimates to be statistically no different from the "true" total. Of course, this conclusion is supported by the higher response rates as well, that are shown in Tables 6.6, 6.7 and 6.8.

6.2 Individual Sample Variance Comparison Using Sum of Weights Adjustment

Efforts will now be made to compare variance estimates obtained from Conditions I and II over the 100 samples of $\hat{V}_1$ to $\hat{V}_2$. There is reason to believe that the median of $\hat{V}_2$ minus $\hat{V}_1$ will be positive, partly because of additional terms in $\hat{V}_2$. Also, thoughts are that Taylor's approximation

TABLE 6.1. COMPARISON OF ESTIMATED TOTALS OVER 100 SAMPLES TO THE "TRUE" TOTAL GENERATED FROM SAMPLE INFORMATION USING SUM OF WEIGHTS ADJUSTMENT

(RESPONSE RATES: 50% & 70%)

Variable	Domain	"True" Total	$\hat{F}_1$
Births	1	1,240,370	1,278,874
	2	1,903,959	1,965,190
	Aggregate	3,144,329	3,244,064
Divorces	1	373,636	389,043
	2	662,022	686,697
	Aggregate	1,035,658	1,075,739
U.S. Population	1	80,302,039	83,217,345
	2	132,763,646	137,025,363
	Aggregate	213,065,685	220,242,709
Food Stores	1	125,547	125,173
	2	148,355	152,190
	Aggregate	273,902	277,363
Gasoline Service Stations	1	113,772	116,588
	2	118,573	122,726
	Aggregate	232,345	239,314
Hospitals	1	3998	4,184
	2	3326	3,422
	Aggregate	7324	7,606
Marriages	1	942,908	949,114
	2	1,208,807	1,217,440
	Aggregate	2,151,715	2,166,554
Public School Enrollment	1	17,757,679	18,385,323
	2	27,031,137	27,816,660
	Aggregate	44,788,816	46,201,983
Vehicle Thefts	1	127,408	129,766
	2	887,005	902,252
	Aggregate	1,014,413	1,032,018

TABLE 6.2. COMPARISON OF ESTIMATED TOTALS OVER 100 SAMPLES TO THE "TRUE" TOTAL GENERATED FROM SAMPLE INFORMATION USING SUM OF WEIGHTS ADJUSTMENT

(RESPONSE RATES: 70% & 90%)

Variable	Domain	"True" Total	$\hat{F}_1$
Births	1	1,240,370	1,227,705
	2	1,903,959	1,927,240
	Aggregate	3,144,329	3,154,945
Divorces	1	373,636	370,109
	2	662,022	670,507
	Aggregate	1,035,658	1,040,616
U.S. Population	1	80,302,039	79,991,742
	2	132,763,646	134,608,390
	Aggregate	213,065,685	214,600,132
Food Stores	1	125,547	123,469
	2	148,355	149,641
	Aggregate	273,902	273,110
Gasoline Service Stations	1	113,772	112,218
	2	118,573	120,606
	Aggregate	232,345	232,824
Hospitals	1	3998	3,987
	2	3326	3,362
	Aggregate	7324	7,349
Marriages	1	942,908	910,310
	2	1,208,807	1,207,862
	Aggregate	2,151,715	2,118,172
Public School Enrollment	1	17,757,679	17,629,721
	2	27,031,137	27,378,185
	Aggregate	44,788,816	45,007,906
Vehicle Thefts	1	127,408	123,870
	2	887,005	887,041
	Aggregate	1,014,413	1,010,911

TABLE 6.3. COMPARISON OF ESTIMATED TOTALS OVER 100 SAMPLES TO THE "TRUE" TOTAL GENERATED FROM SAMPLE INFORMATION USING SUM OF WEIGHTS ADJUSTMENT

(RESPONSE RATES: 90% & 95%)

Variable	Domain	"True" Total	$\hat{F}_1$
Births	1	1,240,370	1,251,074
	2	1,903,959	1,892,874
	Aggregate	3,144,329	3,143,948
Divorces	1	373,636	375,930
	2	662,022	660,672
	Aggregate	1,035,658	1,036,602
U.S. Population	1	80,302,039	81,130,547
	2	132,763,646	132,336,391
	Aggregate	213,065,685	213,466,938
Food Stores	1	125,547	124,281
	2	148,355	146,870
	Aggregate	273,902	271,152
Gasoline Service Stations	1	113,772	113,300
	2	118,573	118,712
	Aggregate	232,345	232,012
Hospitals	1	3998	4,058
	2	3326	3,313
	Aggregate	7324	7,372
Marriages	1	942,908	912,504
	2	1,208,807	1,767,765
	Aggregate	2,151,715	2,089,268
Public School Enrollment	1	17,757,679	17,918,400
	2	27,031,137	26,912,867
	Aggregate	44,788,816	44,831,267
Vehicle Thefts	1	127,408	127,298
	2	887,005	868,672
	Aggregate	1,014,413	995,970

TABLE 6.4. COMPARISON OF ESTIMATED TOTALS
OVER 100 SAMPLES TO THE "TRUE" TOTAL
GENERATED FROM SAMPLE INFORMATION
USING SUM OF WEIGHTS ADJUSTMENT

(RESPONSE RATES: 100% & 100%)

Variable	Domain	"True" Total	$\hat{F}_1$
Births	1	1,240,370	1,239,343
	2	1,903,959	1,902,233
	Aggregate	3,144,329	3,141,575
Divorces	1	373,636	372,992
	2	662,022	663,842
	Aggregate	1,035,658	1,036,834
U.S. Population	1	80,302,039	80,285,454
	2	132,763,646	132,855,004
	Aggregate	213,065,685	213,140,458
Food Stores	1	125,547	123,379
	2	148,355	147,677
	Aggregate	273,902	271,056
Gasoline Service Stations	1	113,772	112,562
	2	118,573	119,120
	Aggregate	232,345	231,683
Hospitals	1	3998	4,044
	2	3326	3,321
	Aggregate	7324	7,366
Marriages	1	942,908	902,393
	2	1,208,807	1,188,486
	Aggregate	2,151,715	2,090,879
Public School Enrollment	1	17,757,679	17,737,508
	2	27,031,137	27,022,286
	Aggregate	44,788,816	44,759,793
Vehicle Thefts	1	127,408	126,167
	2	887,005	873,128
	Aggregate	1,014,413	999,295

TABLE 6.5. COMPARISON OF THE ESTIMATED TOTALS OVER 100 SAMPLES TO THE "TRUE" TOTAL GENERATED FROM SAMPLE INFORMATION USING POPULATION WEIGHTING CLASS ADJUSTMENT

(RESPONSE RATES: 50% & 70%)

Variable	Domain	"True" Total	$\hat{F}_1$
Births	1	1,240,370	1,238,493
	2	1,903,959	1,898,877
	Aggregate	3,144,329	3,137,370
Divorces	1	373,636	376,088
	2	662,022	663,093
	Aggregate	1,035,658	1,039,181
U.S. Population	1	80,302,039	80,553,229
	2	132,763,646	132,334,699
	Aggregate	213,065,685	212,887,927
Food Stores	1	125,547	122,203
	2	148,355	147,112
	Aggregate	273,902	269,315
Gasoline Service Stations	1	113,772	113,598
	2	118,573	118,665
	Aggregate	232,345	232,264
Hospitals	1	3998	4,105
	2	3326	3,312
	Aggregate	7324	7,417
Marriages	1	942,908	925,608
	2	1,208,807	1,176,754
	Aggregate	2,151,715	2,102,362
Public School Enrollment	1	17,757,679	17,772,557
	2	27,031,137	26,883,700
	Aggregate	44,788,816	44,656,257
Vehicle Thefts	1	127,408	126,659
	2	887,005	875,184
	Aggregate	1,014,413	1,001,843

TABLE 6.6. COMPARISON OF THE ESTIMATED TOTALS OVER 100 SAMPLES TO THE "TRUE" TOTAL GENERATED FROM SAMPLE INFORMATION USING POPULATION WEIGHTING CLASS ADJUSTMENT

(RESPONSE RATES: 70% & 90%)

Variable	Domain	"True" Total	$\hat{F}_1$
Births	1	1,240,370	1,227,559
	2	1,903,959	1,908,001
	Aggregate	3,144,329	3,135,560
Divorces	1	373,636	369,969
	2	662,022	663,768
	Aggregate	1,035,658	1,033,736
U.S. Population	1	80,302,039	79,939,942
	2	132,763,646	133,262,809
	Aggregate	213,065,685	213,202,751
Food Stores	1	125,547	123,204
	2	148,355	148,198
	Aggregate	273,902	271,402
Gasoline Service Stations	1	113,772	112,360
	2	118,573	119,400
	Aggregate	232,345	231,760
Hospitals	1	3998	3,987
	2	3326	3,333
	Aggregate	7324	7,320
Marriages	1	942,908	905,987
	2	1,208,807	1,195,178
	Aggregate	2,151,715	2,101,165
Public School Enrollment	1	17,757,679	17,623,612
	2	27,031,137	27,099,556
	Aggregate	44,788,816	44,723,167
Vehicle Thefts	1	127,408	124,485
	2	887,005	877,510
	Aggregate	1,014,413	1,001,996

TABLE 6.7. COMPARISON OF THE ESTIMATED TOTALS OVER 100 SAMPLES TO THE "TRUE" TOTAL GENERATED FROM SAMPLE INFORMATION USING POPULATION WEIGHTING CLASS ADJUSTMENT

(RESPONSE RATES: 90% & 95%)

Variable	Domain	"True" Total	$\hat{F}_1$
Births	1	1,240,370	1,243,659
	2	1,903,959	1,895,446
	Aggregate	3,144,329	3,139,106
Divorces	1	373,636	374,269
	2	662,022	661,011
	Aggregate	1,035,658	1,035,279
U.S. Population	1	80,302,039	80,624,553
	2	132,763,646	132,471,121
	Aggregate	213,065,685	213,095,674
Food Stores	1	125,547	123,274
	2	148,355	147,026
	Aggregate	273,902	270,300
Gasoline Service Stations	1	113,772	112,679
	2	118,573	118,747
	Aggregate	232,345	231,426
Hospitals	1	3998	4,046
	2	3326	3,309
	Aggregate	7324	7,356
Marriages	1	942,908	907,197
	2	1,208,807	1,175,114
	Aggregate	2,151,715	2,082,311
Public School Enrollment	1	17,757,679	17,801,011
	2	27,031,137	26,943,362
	Aggregate	44,788,816	44,744,373
Vehicle Thefts	1	127,408	126,946
	2	887,005	867,469
	Aggregate	1,014,413	994,415

TABLE 6.8. COMPARISON OF THE ESTIMATED TOTALS OVER 100 SAMPLES TO THE "TRUE" TOTAL GENERATED FROM SAMPLE INFORMATION USING POPULATION WEIGHTING CLASS ADJUSTMENT

(RESPONSE RATES: 100% & 100%)

Variable	Domain	"True" Total	$\hat{F}_1$
Births	1	1,240,370	1,239,343
	2	1,903,959	1,902,233
	Aggregate	3,144,329	3,141,575
Divorces	1	373,636	372,992
	2	662,022	663,842
	Aggregate	1,035,658	1,036,834
U.S. Population	1	80,302,039	80,285,454
	2	132,763,646	132,855,004
	Aggregate	213,065,685	213,140,458
Food Stores	1	125,547	123,379
	2	148,355	147,677
	Aggregate	273,902	271,056
Gasoline Service Stations	1	113,772	112,562
	2	118,573	119,120
	Aggregate	232,345	231,683
Hospitals	1	3998	4,044
	2	3326	3,321
	Aggregate	7324	7,366
Marriages	1	942,908	902,393
	2	1,208,807	1,188,486
	Aggregate	2,151,715	2,090,879
Public School Enrollment	1	17,757,679	17,737,508
	2	27,031,137	27,022,286
	Aggregate	44,788,816	44,759,793
Vehicle Thefts	1	127,408	126,167
	2	887,005	873,128
	Aggregate	1,014,413	999,295

yield underestimates with this adjustment. However, the hypothesis is, although $\hat{V}_2$ produces an underestimate, it underestimates less than $\hat{V}_1$.

The Sign Test as described in section 6.1 will be used to test the following hypothesis:

Ho: $\theta = 0$

Ha: $\theta > 0$

where θ = the median of the differences (Z_{ijk}'s),

$Z_{ijk} = \hat{V}_{2ijk} - \hat{V}_{1ijk}$.

The individual variances correspond to the i^{th} sample, j^{th} variable of interest, and k^{th} domain obtained from Conditions II and I, respectively.

Table 6.9 contains the results of the paired Sign Test of $\hat{V}_1$ vs $\hat{V}_2$. Each of the 9 variables for domain #1 proves to be nonsignificant at the low response rate. However, domain #2 yields significance at a level $\leq .01$ consistently for all variables. The aggregate of domains is highly significant at the low rate for all variables except food stores, gasoline service stations, hospitals, and marriages. This nonsignificance may be explained by the fact that county data values for these 4 variables do not vary greatly. On the other hand, county data values for the other 5 variables vary tremendously, especially the county population. $\hat{V}_1$ reflects this variability and underestimates the variance consistently at the low response rate.

The medium response rate indicates fewer significant differences than the lower rate. Domain #2 still shows significant differences at a rate $\leq .01$ for all variables. Fewer differences are discovered for the aggregate.

At the high response rate, the aggregate yields no significant differences and domain #2 indicate 3 variables to be different $\leq .01$

TABLE 6.9. DISTRIBUTION-FREE SIGN TEST
GENERATED FROM SAMPLE INFORMATION
USING SUM OF WEIGHTS ADJUSTMENT

$\hat{V}_2$ vs $\hat{V}_1$

Ho: $\theta = 0$

Ha: $\theta > 0$

Variable	Domain	RESPONSE RATE 50% & 70%	70% & 90%	90% & 95%
Births	1			
	2	**	**	**
	Aggregate	**		
Divorces	1			
	2	**	**	
	Aggregate	**	*	
U.S. Population	1			
	2	**	**	**
	Aggregate	**		
Food Stores	1			
	2	**	**	*
	Aggregate			
Gasoline Service Stations	1			
	2	**	**	
	Aggregate			
Hospitals	1			
	2	**	**	*
	Aggregate			
Marriages	1			
	2	**	**	
	Aggregate			
Public School Enrollment	1			
	2	**	**	**
	Aggregate	**		
Vehicle Thefts	1			
	2	**	**	
	Aggregate			

where: θ = median of the differences (Z_{ijk}'s)

$Z_{ijk} = \hat{V}_{2ijk} - \hat{V}_{1ijk}$, i^{th} sample, j^{th} variable, k^{th} domain

Under Ho: $P(Z_{ijk} > 0) = P(Z_{ijk} < 0) = \frac{1}{2}$

*Significant at the .05 level

**Significant at $\leq$ .01 level

level and 2 variables at the .05 level. Therefore, the Sign Test results appear to be consistent with what is thought to happen as the response rate improves. The better the response, the more similar are $\hat{V}_1$ and $\hat{V}_2$. With a perfect response rate, in theory, $\hat{V}_1$ and $\hat{V}_2$ are equal.

For each response rate the following paired t-test was used to test $\hat{V}_2$ vs $\hat{V}_1$:

Ho: $\mu_d = 0$

Ha: $\mu_d > 0$

where μ_d is the mean of the population of differences

$d_{ijk} = \hat{S}_{2ijk} - \hat{S}_{1ijk}$.

The individual standard deviations (square root of the respective variances) correspond to sample i, variable j, and domain k. In general, the paired t-test conclusions were similar.

6.3 Individual Sample Variance Comparison Using Population Weighting Class Adjustment

Due to the type of adjustment made by using the county population, the belief is that the median of $\hat{V}_1$ will be larger than that of $\hat{V}_2$. The null hypothesis is then:

Ho: $\theta = 0$

Ha: $\theta > 0$

where: θ = median of the differences (Z_{ijk}'s)

$Z_{ijk} = \hat{V}_{1ijk} - \hat{V}_{2ijk}$

This type of adjustment indicates differences which are highly significant at low and high response rates for all variables for both domains and the aggregate. The high response rate tends to yield fewer significant differences. Results are in Table 6.10.

Again, the paired t-test in general substantiates these conclusions.

TABLE 6.10. DISTRIBUTION-FREE SIGN TEST
GENERATED FROM SAMPLE INFORMATION
USING POPULATION WEIGHTING CLASS ADJUSTMENT

$\hat{V}_2$ vs $\hat{V}_1$

Ho: $\theta = 0$

Ha: $\theta > 0$

Variable	Domain	RESPONSE RATE 50% & 70%	70% & 90%	90% & 95%
Births	1	**	**	*
	2	**	**	**
	Aggregate	**	**	**
Divorces	1	**	**	*
	2	**	**	**
	Aggregate	**	**	**
U.S. Population	1	**	**	**
	2	**	**	**
	Aggregate	**	**	**
Food Stores	1	**	**	*
	2	**	**	**
	Aggregate	**	**	**
Gasoline Service	1	**	**	*
Stations	2	**	**	
	Aggregate	**	**	**
Hospitals	1	**	**	
	2	**	**	*
	Aggregate	**	**	**
Marriages	1	**	**	**
	2	**	**	**
	Aggregate	**	**	**
Public School	1	**	**	**
Enrollment	2	**	**	*
	Aggregate	**	**	**
Vehicle Thefts	1	**	**	
	2	**	**	**
	Aggregate	**	**	**

where: θ = median of the differences (Z_{ijk}'s)

$Z_{ijk} = \hat{V}_{2ijk} - \hat{V}_{1ijk}$, i^{th} sample, j^{th} variable, k^{th} domain

Under Ho: $P(Z_{ijk} > 0) = P(Z_{ijk} < 0) = \frac{1}{2}$

*Significant at the .05 level
**Significant at $\leq$ .01 level

6.4 Average Variance Comparison Over 100 Samples

An unbiased replicated estimator of the variance of $\hat{F}_1$ will now be compared to $\hat{V}_1$ and $\hat{V}_2$. The replicated variance estimator will be denoted as $\hat{V}$ where:

$$\hat{V} = \sum_{i=1}^{100} \left(\frac{\hat{F} - \bar{F}}{n - 1}\right)^2 \quad .$$

where $\hat{F} = \hat{F}_1$, $\bar{F} = \bar{F}_1$, since t_{1c} is generated from sample information. As previously stated, $\hat{V}$ is known to be an unbiased estimator but it is also a consistent estimator. Of the two estimators, $\hat{V}_1$ and $\hat{V}_2$, the one that is more stable and comes closer to $\hat{V}$ will be considered the better estimator.

The process will involve comparing $\hat{V}$ to:

i) the mean of $\hat{V}_1$ denoted by $\bar{V}_1$

ii) the mean of $\hat{V}_2$ denoted by $\bar{V}_2$.

It follows, for any variable and domain, that:

$$\bar{V}_1 = \frac{\sum_{i=1}^{100} \hat{V}_{1i}}{100} \quad , \quad \bar{V}_2 = \frac{\sum_{i=1}^{100} \hat{V}_{2i}}{100} \quad .$$

The above notation will be used for both types of weighting class adjustments.

6.4.1 Using Sum of Weights Adjustment

From Table 6.11, the low response rate table, it is clear that $\bar{V}_1$ appears to vastly underestimate the variance of $\hat{F}_1$ for domain #2 and also for the aggregate. This underestimation is reflected for all 9 variables at the low response rate. One reason that could be considered as having contributed to these underestimates is due to the adjustment factor

TABLE 6.11. COMPARISON OF VARIANCE ESTIMATES
OVER 100 SAMPLES
GENERATED FROM SAMPLE INFORMATION
USING SUM OF WEIGHTS ADJUSTMENT

(RESPONSE RATES: 50% & 70%)

Variable	Domain	$\hat{V}$	$\bar{V}_1$	$\bar{V}_2$
Births	1	9.366 E+10	1.090 E+11	1.090 E+11
	2	3.400 E+11	8.290 E+10	2.465 E+11
	Aggregate	3.605 E+11	1.030 E+11	3.070 E+11
Divorces	1	1.155 E+10	1.243 E+10	1.191 E+10
	2	4.611 E+10	1.371 E+10	3.184 E+10
	Aggregate	5.203 E+10	1.687 E+10	3.935 E+10
U.S. Population	1	3.931 E+14	4.472 E+14	4.374 E+14
	2	1.710 E+15	3.761 E+14	1.174 E+15
	Aggregate	1.722 E+15	4.207 E+14	1.359 E+15
Food Stores	1	1,018,163,445	1,262,901,550	1,106,869,201
	2	2,123,496,814	501,270,573	1,458,360,805
	Aggregate	2,385,214,728	1,125,875,623	2,094,742,370
Gasoline Service Stations	1	823,899,844	980,513,576	883,737,057
	2	1,301,606,508	378,268,414	947,397,310
	Aggregate	1,568,707,408	878,345,972	1,522,700,106
Hospitals	1	1,711,620	1,941,797	1,633,909
	2	1,028,275	295,709	751,430
	Aggregate	2,473,985	1,752,915	1,994,747
Marriages	1	1.119 E+11	1.306 E+11	1.068 E+11
	2	1.511 E+11	5.059 E+10	1.088 E+11
	Aggregate	2.399 E+11	1.548 E+11	2.028 E+11
Public School Enrollment	1	1.978 E+13	2.191 E+13	2.187 E+13
	2	6.850 E+13	1.654 E+13	4.950 E+13
	Aggregate	7.209 E+13	2.066 E+13	6.080 E+13
Vehicle Thefts	1	1,728,812,251	1,659,530,926	1,679,335,495
	2	1.145 E+11	4.177 E+10	7.839 E+10
	Aggregate	1.120 E+11	3.935 E+10	7.831 E+10

t_{1c}/t_{2c} being greater than 2.0. The adjustment is greater than 2.0 on occasions only for response rates of 50% and 70%. Chapman indicated that at a low reponse rate one might have to settle for a larger adjustment factor.

According to Table 6.11, both $\bar{V}_1$ and $\bar{V}_2$ overestimate the replicated variance for most variables in domain #1. However, $\bar{V}_2$ is always closer to $\hat{V}$. The reason why both estimators generate over estimates is primarily due to the variation of county data values within domains. There is a strong positive correlation between increase in population and increase in data values for each variable. Domain #1 includes only those counties whose 1970 population is less than 135,000, a range of 129,000. Domain #2 consists of those with populations ranging from approximately 135,000 to 7,000,000, a range of 6,865,000. Population ranges differ considerably between domains. It is then feasible to assume that for each variable, the range of data values between domains differ in a similar manner. This difference is reflected in the variance estimates.

The mean estimator $\bar{V}_2$, underestimates the variance of $\hat{F}_1$ for domain #2 and the aggregate. However, observing Table 6.11, $\bar{V}_2$ comes closer to the replicated estimate than $\bar{V}_1$ does. Inclusion of the additional terms has thus yielded better estimates.

Overall, the estimates obtained with the medium response rate are better than those from the low response rate. They are closer to the replicated estimate. This fact holds for both $\bar{V}_1$ and $\bar{V}_2$ which is shown in Table 6.12. The average variance estimates are smaller for the medium rate than for the low rate. The high response rate produced

TABLE 6.12. COMPARISON OF VARIANCE ESTIMATES
OVER 100 SAMPLES
GENERATED FROM SAMPLE INFORMATION
USING SUM OF WEIGHTS ADJUSTMENT

(RESPONSE RATES: 70% & 90%)

Variable	Domain	$\hat{V}$	$\bar{V}_1$	$\bar{V}_2$
Births	1	8.024 E+10	7.113 E+10	8.202 E+10
	2	1.054 E+11	5.625 E+10	1.012 E+11
	Aggregate	8.056 E+10	4.054 E+10	9.881 E+10
Divorces	1	9,184,298,042	8,413,478,500	9,097,719,857
	2	1.576 E+10	9,398,189,942	1.498 E+10
	Aggregate	1.416 E+10	8,921,641,372	1.561 E+10
U.S. Population	1	3.213 E+14	2.863 E+14	3.336 E+14
	2	4.926 E+14	2.504 E+14	4.526 E+14
	Aggregate	3.484 E+14	1.425 E+14	4.041 E+14
Food Stores	1	1,116,173,793	898,499,904	1,017,835,735
	2	645,811,446	342,404,725	592,743,765
	Aggregate	1,026,661,612	556,724,419	940,366,628
Gasoline Service Stations	1	713,158,805	661,196,721	725,979,802
	2	427,011,185	251,999,967	424,104,203
	Aggregate	636,118,219	445,324,024	701,329,612
Hospitals	1	1,769,367	1,430,207	1,475,690
	2	353,331	201,902	324,050
	Aggregate	1,572,040	1,138,749	1,326,537
Marriages	1	1.069 E+11	9.813 E+10	9.776 E+10
	2	7.742 E+10	5.567 E+10	6.958 E+10
	Aggregate	1.856 E+11	1.528 E+11	1.641 E+11
Public School Enrollment	1	1.539 E+13	1.407 E+13	1.640 E+13
	2	2.089 E+13	1.122 E+13	1.971 E+13
	Aggregate	1.536 E+13	7.901 E+13	1.940 E+13
Vehicle Thefts	1	1,104,346,616	1,181,901,545	1,209,415,432
	2	5.537 E+10	3.010 E+10	3.888 E+10
	Aggregate	5.030 E+10	2.745 E+10	3.650 E+10

even smaller variances which is evident in Table 6.13. But, the ratios $\hat{R}_1 = (\bar{V}_1/\hat{V}) \times 100$ and $\hat{R}_2 = (\bar{V}_2/\hat{V}) \times 100$ from Table 6.14 are quite similar to those at the medium rate. Observing the estimates obtained from $\hat{V}_1$ and $\hat{V}_2$ over different response rates, it is clear that $\hat{V}_2$ is more stable than $\hat{V}_1$.

At the Perfect Response Rate, from Table 6.15, $\bar{V}_1$ equals $\bar{V}_2$ and both estimators generate "good" estimates of the variance. The ratios in Table 6.14, in general, are not far from 100%.

6.4.2 Using Population Weighting Class Adjustment

The researcher should expect a smaller estimate of variance if the population weighting class adjustment is used. This particular procedure tends to stabilize adjustments between weighting classes and, therefore, reduces the variance.

Since $\hat{V}_1$ is thought to be larger than $\hat{V}_2$, it will also follow that $\bar{V}_1$ will be larger than $\bar{V}_2$. This can be seen from the tables of average estimates, Tables 6.16, 6.17, 6.18, 6.19, and more clearly from the ratios in Table 6.20. $\bar{V}_1$ reacts to the variability of data for the aggregate moreso than $\bar{V}_2$ does. For six variables, $\bar{V}_1$ vastly overestimates the aggregate. The variables are births, divorces, food stores, gasoline service stations, public school enrollment, and U.S. population. However, $\bar{V}_2$ yields reasonable estimates for these six variables which may be some indication that $\hat{V}_2$ is a better estimator than $\hat{V}_1$.

TABLE 6.13. COMPARISON OF VARIANCE ESTIMATES
OVER 100 SAMPLES
GENERATED FROM SAMPLE INFORMATION
USING SUM OF WEIGHTS ADJUSTMENT

(RESPONSE RATES: 90% & 95%)

Variable	Domain	$\hat{V}$	$\bar{V}_1$	$\bar{V}_2$
Births	1	5.988 E+10	6.129 E+10	6.185 E+10
	2	7.387 E+10	4.863 E+10	6.689 E+10
	Aggregate	5.412 E+10	2.130 E+10	4.627 E+10
Divorces	1	6,738,828,061	7,071,387,773	7,212,674,730
	2	1.255 E+10	8,385,698,034	1.133 E+10
	Aggregate	1.133 E+10	6,508,391,707	1.038 E+10
U.S. Population	1	2.354 E+14	2.322 E+14	2.382 E+14
	2	3.615 E+14	2.187 E+14	3.102 E+14
	Aggregate	2.325 E+14	5.458 E+14	1.802 E+14
Food Stores	1	739,235,927	739,415,393	753,561,828
	2	491,102,676	297,895,903	409,657,210
	Aggregate	614,519,420	332,628,363	511,910,953
Gasoline Service Stations	1	521,992,265	540,267,401	560,174,300
	2	362,608,980	224,164,637	300,344,141
	Aggregate	459,725,518	278,120,062	423,919,381
Hospitals	1	1,088,737	1,158,371	1,177,452
	2	353,933	181,946	256,251
	Aggregate	1,026,146	850,200	1,018,461
Marriages	1	6.796 E+10	6.558 E+10	6.553 E+10
	2	6.174 E+10	3.689 E+10	6.550 E+10
	Aggregate	1.202 E+11	7.815 E+10	1.129 E+11
Public School Enrollment	1	1.163 E+13	1.135 E+13	1.153 E+13
	2	1.481 E+13	9.669 E+12	1.333 E+13
	Aggregate	1.063 E+13	3.447 E+12	8.481 E+12
Vehicle Thefts	1	1,038,799,723	1,062,607,735	1,050,337,494
	2	5.013 E+10	1.966 E+10	3.449 E+10
	Aggregate	4.620 E+10	2.476 E+10	3.185 E+10

TABLE 6.14. PERCENT OF VARIANCE ESTIMATES TO UNBIASED ESTIMATES GENERATED FROM SAMPLE INFORMATION USING SUM OF WEIGHTS ADJUSTMENT

		RESPONSE RATES						
		50% & 70%		70% & 90%		90% & 95%		100% & 100%
Variable	Domain	$\bar{R}_1$	$\bar{R}_2$	$\bar{R}_1$	$\bar{R}_2$	$\bar{R}_1$	$\bar{R}_2$	$\bar{R}_1 = \bar{R}_2$
Births	1	116.4	116.3	88.6	102.2	102.4	103.3	100.7
	2	24.4	72.5	53.4	96.0	65.8	90.6	102.8
	Aggregate	28.6	85.2	50.3	122.7	39.4	85.5	94.6
Divorces	1	107.6	103.1	91.6	99.1	104.9	107.0	102.3
	2	29.7	69.1	59.6	95.1	66.8	90.3	101.3
	Aggregate	32.4	75.6	63.0	110.2	57.4	91.6	98.3
U.S. Population	1	113.8	111.3	89.1	103.8	98.6	101.1	99.1
	2	22.0	68.7	50.8	91.9	60.5	85.8	102.4
	Aggregate	24.4	78.9	40.9	116.0	23.5	77.5	101.5
Food Stores	1	124.0	108.7	80.5	91.2	100.0	101.2	104.6
	2	23.6	68.7	53.0	91.8	60.7	83.4	92.8
	Aggregate	47.2	87.8	54.2	91.6	54.1	83.3	104.4
Gasoline Service Stations	1	119.0	107.3	92.7	101.8	103.5	107.3	101.1
	2	29.0	72.7	59.0	99.3	61.8	82.8	103.3
	Aggregate	56.0	97.1	70.0	110.2	60.5	92.2	96.9
Hospitals	1	113.4	95.4	80.8	83.4	106.4	108.1	106.5
	2	28.8	73.1	57.1	91.7	51.4	72.4	95.4
	Aggregate	70.9	80.6	72.5	84.4	82.9	99.2	99.3
Marriages	1	116.7	95.4	91.8	91.4	96.5	96.4	92.6
	2	33.5	72.0	71.9	89.9	59.8	106.1	95.1
	Aggregate	64.5	84.5	82.3	88.4	65.0	93.9	88.2
Public School Enrollment	1	110.8	110.6	91.4	106.6	97.6	99.1	95.9
	2	24.1	72.3	53.7	94.3	65.3	90.0	106.3
	Aggregate	28.7	84.3	51.4	126.3	32.4	79.8	97.3
Vehicle Thefts	1	96.0	97.1	107.0	109.5	102.3	101.1	99.8
	2	36.5	68.5	54.4	70.2	55.3	68.8	75.7
	Aggregate	35.1	69.9	54.6	72.6	53.6	68.9	75.5

$\bar{R}_i = \frac{\bar{V}_i}{\hat{V}}$ (for i=1 or 2)

TABLE 6.15. COMPARISON OF VARIANCE ESTIMATES OVER 100 SAMPLES GENERATED FROM SAMPLE INFORMATION USING SUM OF WEIGHTS ADJUSTMENT

(RESPONSE RATES: 100% & 100%)

Variable	Domain	$\hat{V}$	$\bar{V}_1 = \bar{V}_2$
Births	1	5.467 E+10	5.507 E+10
	2	4.443 E+10	4.567 E+10
	Aggregate	1.048 E+10	9,912,354,218
Divorces	1	6,514,425,447	6,662,484,588
	2	7,862,499,812	7,964,524,960
	Aggregate	5,615,918,454	5,520,273,769
U.S. Population	1	2.099 E+14	2.080 E+14
	2	1.987 E+14	2.034 E+14
	Aggregate	6.726 E+12	6.827 E+12
Food Stores	1	639,289,700	668,595,587
	2	298,819,933	277,414,586
	Aggregate	223,900,018	233,758,373
Gasoline Service Stations	1	492,235,308	497,540,789
	2	203,291,963	210,091,568
	Aggregate	226,107,965	219,188,375
Hospitals	1	1,058,956	1,127,543
	2	176,419	168,244
	Aggregate	799,301	794,042
Marriages	1	6.283 E+10	5.819 E+10
	2	4.742 E+10	4.511 E+10
	Aggregate	1.022 E+11	9.013 E+10
Public School Enrollment	1	1.047 E+13	1.004 E+13
	2	8.563 E+12	9.100 E+12
	Aggregate	1.228 E+12	1.195 E+12
Vehicle Thefts	1	970,454,892	968,768,358
	2	3.486 E+10	2.640 E+10
	Aggregate	3.105 E+10	2.343 E+10

TABLE 6.16. COMPARISON OF VARIANCE ESTIMATES OVER 100 SAMPLES GENERATED FROM SAMPLE INFORMATION USING POPULATION WEIGHTING CLASS ADJUSTMENT

(RESPONSE RATES: 50% & 70%)

Variable	Domain	$\hat{V}$	$\bar{V}_1$	$\bar{V}_2$
Births	1	6.506 E+10	9.582 E+10	6.456 E+10
	2	5.621 E+10	7.694 E+10	5.343 E+10
	Aggregate	1.432 E+10	8.929 E+10	1.273 E+10
Divorces	1	8.368 E+9	1.056 E+10	7.803 E+10
	2	8.720 E+9	1.240 E+10	9.310 E+10
	Aggregate	6.270 E+9	1.457 E+10	6.371 E+10
U.S. Population	1	2.656 E+14	3.880 E+14	2.448 E+14
	2	2.461 E+14	3.444 E+14	2.369 E+14
	Aggregate	9.610 E+12	3.584 E+14	8.399 E+12
Food Stores	1	8.731 E+8	1.161 E+9	7.664 E+8
	2	3.954 E+8	4.540 E+8	3.236 E+8
	Aggregate	3.206 E+8	1.006 E+9	3.079 E+8
Gasoline Service Stations	1	7.047 E+8	8.893 E+8	6.100 E+8
	2	2.509 E+8	3.390 E+8	2.477 E+8
	Aggregate	3.189 E+8	7.778 E+8	2.852 E+8
Hospitals	1	1.534 E+6	1.889 E+6	1.464 E+6
	2	2.274 E+5	2.680 E+5	1.464 E+%
	Aggregate	1.189 E+6	1.702 E+6	1.094 E+6
Marriages	1	1.052 E+11	1.283 E+11	9.845 E+10
	2	4.491 E+10	5.145 E+10	4.164 E+10
	Aggregate	1.144 E+11	1.597 E+11	1.096 E+11
Public School Enrollment	1	1.289 E+13	1.899 E+13	1.185 E+13
	2	1.025 E+13	1.514 E+13	1.074 E+13
	Aggregate	1.934 E+12	1.761 E+13	1.651 E+12
Vehicle Thefts	1	1.564 E+9	1.515 E+9	1.278 E+9
	2	4.251 E+10	3.742 E+10	3.263 E+10
	Aggregate	3.805 E+10	3.509 E+10	2.915 E+10

TABLE 6.17. COMPARISON OF VARIANCE ESTIMATES OVER 100 SAMPLES GENERATED FROM SAMPLE INFORMATION USING POPULATION WEIGHTING CLASS ADJUSTMENT

(RESPONSE RATES: 70% & 90%)

Variable	Domain	$\hat{V}$	$\bar{V}_1$	$\bar{V}_2$
Births	1	6.191 E+10	6.886 E+10	5.673 E+10
	2	4.937 E+10	5.497 E+10	4.817 E+10
	Aggregate	1.154 E+10	3.760 E+10	1.130 E+10
Divorces	1	7.816 E+9	8.195 E+9	7.174 E+9
	2	8.655 E+9	9.142 E+9	8.436 E+9
	Aggregate	6.328 E+9	8.420 E+9	5.982 E+9
U.S. Population	1	2.315 E+14	2.758 E+14	2.184 E+14
	2	2.164 E+14	2.448 E+14	2.135 E+14
	Aggregate	7.931 E+12	1.281 E+14	7.880 E+12
Food Stores	1	7.590 E+8	8.530 E+8	7.001 E+8
	2	3.307 E+8	3.323 E+8	2.920 E+8
	Aggregate	2.722 E+8	5.041 E+8	2.596 E+8
Gasoline Service Stations	1	5.498 E+8	6.486 E+8	5.185 E+8
	2	2.197 E+8	2.452 E+8	2.214 E+8
	Aggregate	2.353 E+8	4.291 E+8	2.303 E+8
Hospitals	1	1.492 E+6	1.355 E+6	1.192 E+6
	2	2.022 E+5	1.979 E+5	1.790 E+5
	Aggregate	1.140 E+6	1.070 E+6	8.608 E+5
Marriages	1	8.965 E+10	8.731 E+10	7.653 E+10
	2	5.345 E+10	5.379 E+10	5.123 E+10
	Aggregate	1.387 E+11	1.336 E+10	1.164 E+11
Public School Enrollment	1	1.148 E+13	1.346 E+13	1.046 E+13
	2	9.183 E+12	1.094 E+13	9.586 E+12
	Aggregate	1.453 E+12	7.102 E+12	1.340 E+12
Vehicle Thefts	1	1.163 E+9	1.163 E+9	1.061 E+9
	2	4.044 E+10	2.981 E+10	2.870 E+10
	Aggregate	3.631 E+10	2.720 E+10	2.574 E+10

TABLE 6.18. COMPARISON OF VARIANCE ESTIMATES OVER 100 SAMPLES GENERATED FROM SAMPLE INFORMATION USING POPULATION WEIGHTING CLASS ADJUSTMENT

(RESPONSE RATES: 90% & 95%)

Variable	Domain	$\hat{V}$	$\bar{V}_1$	$\bar{V}_2$
Births	1	5.567 E+10	5.965 E+10	5.545 E+10
	2	4.403 E+10	4.815 E+10	4.578 E+10
	Aggregate	1.127 E+10	1.994 E+10	1.014 E+10
Divorces	1	6.605 E+9	7.003 E+9	6.686 E+9
	2	7.704 E+9	8.174 E+9	7.890 E+9
	Aggregate	5.394 E+9	6.256 E+9	5.430 E+9
U.S. Population	1	2.141 E+14	2.267 E+14	2.086 E+14
	2	2.013 E+14	2.160 E+14	2.045 E+14
	Aggregate	7.137 E+12	5.010 E+13	6.934 E+12
Food Stores	1	6.530 E+8	7.105 E+8	6.676 E+8
	2	2.976 E+8	2.934 E+8	2.780 E+8
	Aggregate	2.381 E+8	3.132 E+8	2.390 E+8
Gasoline Service Stations	1	5.118 E+8	5.314 E+8	4.994 E+8
	2	2.090 E+8	2.202 E+8	2.112 E+8
	Aggregate	2.320 E+8	2.723 E+8	2.226 E+8
Hospitals	1	1.161 E+6	1.152 E+6	1.117 E+6
	2	1.776 E+5	1.772 E+5	1.698 E+5
	Aggregate	8.838 E+5	8.466 E+5	7.979 E+5
Marriages	1	6.698 E+10	6.400 E+10	6.173 E+10
	2	3.386 E+10	3.325 E+10	3.341 E+10
	Aggregate	8.123 E+10	7.374 E+10	7.102 E+10
Public School Enrollment	1	1.050 E+13	1.102 E+13	1.010 E+13
	2	8.539 E+12	9.590 E+12	9.161 E+12
	Aggregate	1.186 E+12	3.201 E+12	1.229 E+12
Vehicle Thefts	1	1.041 E+9	1.063 E+9	1.021 E+9
	2	3.430 E+10	2.683 E+10	2.606 E+10
	Aggregate	3.001 E+10	2.390 E+10	2.296 E+10

TABLE 6.19. COMPARISON OF VARIANCE ESTIMATES OVER 100 SAMPLES GENERATED FROM SAMPLE INFORMATION USING POPULATION WEIGHTING CLASS ADJUSTMENT

(RESPONSE RATES: 100% & 100%)

Variable	Domain	$\hat{V}$	$\bar{V}_1 = \bar{V}_2$
Births	1	5.466 E+10	5.507 E+10
	2	4.442 E+10	4.567 E+10
	Aggregate	1.047 E+10	9.912 E+9
Divorces	1	6.514 E+9	6.662 E+9
	2	7.862 E+9	7.964 E+9
	Aggregate	5.615 E+9	5.520 E+9
U.S. Population	1	2.099 E+14	2.079 E+14
	2	1.986 E+14	2.034 E+14
	Aggregate	6.725 E+12	6.826 E+12
Food Stores	1	6.329 E+8	6.685 E+8
	2	2.988 E+8	2.774 E+8
	Aggregate	2.239 E+8	2.337 E+8
Gasoline Service Stations	1	4.922 E+8	4.975 E+8
	2	2.032 E+8	2.100 E+8
	Aggregate	2.261 E+8	2.191 E+8
Hospitals	1	1.058 E+6	1.127 E+6
	2	1.764 E+5	1.682 E+5
	Aggregate	7.993 E+5	7.940 E+5
Marriages	1	6.283 E+10	5.819 E+10
	2	4.741 E+10	4.511 E+10
	Aggregate	1.021 E+11	9.012 E+10
Public School Enrollment	1	1.047 E+13	1.003 E+13
	2	8.562 E+12	9.099 E+12
	Aggregate	1.228 E+12	1.195 E+12
Vehicle Thefts	1	9.704 E+9	9.687 E+9
	2	3.486 E+10	2.640 E+10
	Aggregate	3.105 E+10	2.343 E+10

TABLE 6.20. PERCENT OF VARIANCE ESTIMATES TO UNBIASED ESTIMATES GENERATED FROM SAMPLE INFORMATION USING POPULATION WEIGHTING CLASS ADJUSTMENT

		RESPONSE RATES						
		50% & 70%		70% & 90%		90% & 95%		100% & 100%
Variable	Domain	$\hat{R}_1$	$\hat{R}_2$	$\hat{R}_1$	$\hat{R}_2$	$\hat{R}_1$	$\hat{R}_2$	$\hat{R}_1 = \hat{R}_2$
Births	1	147.3	99.2	111.2	91.6	107.2	99.6	100.7
	2	136.9	95.1	111.4	97.6	109.4	104.0	102.8
	Aggregate	623.6	88.9	325.7	97.9	176.9	90.0	94.6
Divorces	1	126.2	93.2	104.9	91.8	106.0	101.2	102.3
	2	142.3	106.8	105.6	97.5	106.1	102.4	101.3
	Aggregate	232.5	101.6	133.1	94.5	116.0	100.7	98.3
U.S. Population	1	146.1	92.2	119.2	94.3	105.9	97.4	99.1
	2	140.0	96.3	113.1	98.7	107.3	101.6	102.4
	Aggregate	3730.3	87.4	1616.3	99.4	702.0	97.2	101.5
Food Stores	1	133.0	87.8	112.4	92.2	108.8	102.2	104.6
	2	114.8	81.8	100.5	88.3	98.6	93.4	92.8
	Aggregate	313.8	96.0	185.2	95.4	131.6	100.4	104.4
Gasoline Service Stations	1	126.2	86.6	118.0	94.3	103.8	97.6	101.1
	2	135.1	98.7	111.6	100.8	105.3	101.0	103.3
	Aggregate	243.9	89.4	182.4	97.9	117.4	96.0	96.9
Hospitals	1	123.1	95.4	90.8	79.9	99.2	96.2	106.5
	2	117.8	88.0	97.9	88.5	99.8	95.6	95.4
	Aggregate	143.2	92.1	93.9	75.5	95.8	90.3	99.3
Marriages	1	122.0	93.6	97.4	85.4	95.6	92.2	92.6
	2	114.6	92.7	100.6	95.8	98.2	98.7	95.1
	Aggregate	139.6	95.8	96.3	83.9	90.8	87.4	88.2
Public School Enrollment	1	147.3	91.9	117.3	91.2	105.0	96.2	95.8
	2	147.7	104.8	119.2	104.4	112.3	107.3	106.3
	Aggregate	910.5	85.4	488.5	92.2	269.9	103.7	97.3
Vehicle Thefts	1	96.8	81.7	100.0	91.2	102.1	98.1	99.8
	2	88.0	76.8	73.7	71.0	78.2	76.0	75.7
	Aggregate	92.2	76.6	74.9	71.0	79.6	76.5	75.5

7. RESULTS FROM EXTERNAL INFORMATION

As previously stated, the number of eligibles in weighting classes of interest are often taken from a census or much larger survey. These totals are considered known and the variance estimates are generally accepted as more reliable and efficient than if they were generated from the current sample information. However, the findings of this survey prove differently. The variance of estimates that were generated from sample information have better properties. The reasons are not fully understood. One justification may be attributed to the relationship between the number of eligibles and the number of respondents in each weighting class. There is a positive correlation between the two estimators. Hence, the ratio, t_{1c}/t_{2c} will generally be smaller than T_{1c}/t_{2c}. The variance estimators based on sample results are formulated to account for this relationship. But estimators $\hat{V}_3$ and $\hat{V}_4$, formulated from Conditions III and IV, respectively, contains an adjustment that in many cases over adjusts for nonresponse and is reflected in the estimated variance. It should be noted that $\hat{V}_1$ and $\hat{V}_4$ estimates are similar in magnitude. Whereas, $\hat{V}_2$ and $\hat{V}_3$ estimates are similar. Therefore, very little could be gained in further displaying tables regarding $\hat{V}_3$ and $\hat{V}_4$. Tables are presented for totals and sign test results, namely Tables 7.1 - 7.8, Tables 7.9, 7.10, respectively.

TABLE 7.1. COMPARISON OF THE ESTIMATES TO THE "TRUE" TOTAL OVER 100 SAMPLES BASED ON EXTERNAL INFORMATION USING SUM OF WEIGHTS ADJUSTMENT

(RESPONSE RATES: 50% & 70%)

Variable	Domain	"True" Total	$\hat{F}_2$
Births	1	1,240,370	1,356,169
	2	1,903,959	2,096,951
	Aggregate	3,144,329	3,453,120
Divorces	1	373,636	406,590
	2	662,022	729,761
	Aggregate	1,035,658	1,136,352
U.S. Population	1	80,302,039	87,919,975
	2	132,763,646	146,281,354
	Aggregate	213,065,685	234,201,328
Food Stores	1	125,547	131,069
	2	148,355	162,860
	Aggregate	273,902	293,930
Gasoline Service Stations	1	113,772	121,422
	2	118,573	130,504
	Aggregate	232,345	251,925
Hospitals	1	3998	4,372
	2	3326	3,642
	Aggregate	7324	8,014
Marriages	1	942,908	1,007,640
	2	1,208,807	1,295,431
	Aggregate	2,151,715	2,303,072
Public School Enrollment	1	17,757,679	19,467,661
	2	27,031,137	29,693,874
	Aggregate	44,788,816	49,161,535
Vehicle Thefts	1	127,408	134,891
	2	887,005	970,896
	Aggregate	1,014,413	1,105,787

TABLE 7.2. COMPARISON OF THE ESTIMATES TO THE "TRUE" TOTAL
OVER 100 SAMPLES
BASED ON EXTERNAL INFORMATION
USING SUM OF WEIGHTS ADJUSTMENT

(RESPONSE RATES: 70% & 90%)

Variable	Domain	"True" Total	$\hat{F}_2$
Births	1	1,240,370	1,288,576
	2	1,903,959	2,094,254
	Aggregate	3,144,329	3,382,829
Divorces	1	373,636	387,241
	2	662,022	726,567
	Aggregate	1,035,658	1,113,808
U.S. Population	1	80,302,039	83,995,736
	2	132,763,646	146,400,246
	Aggregate	213,065,685	230,395,982
Food Stores	1	125,547	128,591
	2	148,355	163,106
	Aggregate	273,902	291,697
Gasoline Service Stations	1	113,772	116,397
	2	118,573	130,600
	Aggregate	232,345	246,996
Hospitals	1	3998	4,096
	2	3326	3,654
	Aggregate	7324	7,750
Marriages	1	942,908	957,830
	2	1,208,807	1,306,969
	Aggregate	2,151,715	2,264,799
Public School Enrollment	1	17,757,679	18,535,525
	2	27,031,137	29,777,365
	Aggregate	44,788,816	48,312,890
Vehicle Thefts	1	127,408	128,757
	2	887,005	972,533
	Aggregate	1,014,413	1,101,290

TABLE 7.3. COMPARISON OF THE ESTIMATES TO THE "TRUE" TOTAL
OVER 100 SAMPLES
BASED ON EXTERNAL INFORMATION
USING SUM OF WEIGHTS ADJUSTMENT

(RESPONSE RATES: 90% & 95%)

Variable	Domain	"True" Total	$\hat{F}_2$
Births	1	1,240,370	1,321,812
	2	1,903,959	2,049,072
	Aggregate	3,144,329	3,370,884
Divorces	1	373,636	395,899
	2	662,022	712,861
	Aggregate	1,035,658	1,108,760
U.S. Population	1	80,302,039	85,696,949
	2	132,763,646	143,419,063
	Aggregate	213,065,685	229,116,012
Food Stores	1	125,547	130,452
	2	148,355	159,424
	Aggregate	273,902	289,876
Gasoline Service	1	113,772	117,981
Stations	2	118,573	128,114
	Aggregate	232,345	246,095
Hospitals	1	3998	4,203
	2	3326	3,583
	Aggregate	7324	7,786
Marriages	1	942,908	964,436
	2	1,208,807	1,269,063
	Aggregate	2,151,715	2,223,499
Public School	1	17,757,679	18,954,829
Enrollment	2	27,031,137	29,174,162
	Aggregate	44,788,816	48,128,991
Vehicle	1	127,408	133,112
Thefts	2	887,005	948,225
	Aggregate	1,014,413	1,081,337

TABLE 7.4. COMPARISON OF THE ESTIMATES TO THE "TRUE" TOTAL
OVER 100 SAMPLES
BASED ON EXTERNAL INFORMATION
USING SUM OF WEIGHTS ADJUSTMENT

(RESPONSE RATES: 100% & 100%)

Variable	Domain	"True" Total	$\hat{F}_2$
Births	1	1,240,370	1,305,296
	2	1,903,959	2,057,499
	Aggregate	3,144,329	3,362,795
Divorces	1	373,636	391,520
	2	662,022	715,624
	Aggregate	1,035,658	1,107,144
U.S. Population	1	80,302,039	84,581,607
	2	132,763,646	143,802,656
	Aggregate	213,065,685	228,384,264
Food Stores	1	125,547	129,258
	2	148,355	160,120
	Aggregate	273,902	289,378
Gasoline Service Stations	1	113,772	116,976
	2	118,573	128,373
	Aggregate	232,345	245,349
Hospitals	1	3998	4,190
	2	3326	3,587
	Aggregate	7324	7,776
Marriages	1	942,908	951,892
	2	1,208,807	1,280,394
	Aggregate	2,151,715	2,232,286
Public School Enrollment	1	17,757,679	18,703,370
	2	27,031,137	29,253,028
	Aggregate	44,788,816	47,956,398
Vehicle Thefts	1	127,408	131,566
	2	887,005	951,430
	Aggregate	1,014,413	1,082,996

TABLE 7.5. COMPARISON OF THE ESTIMATED TOTALS TO THE "TRUE" TOTALS BASED ON EXTERNAL INFORMATION USING POPULATION WEIGHTING CLASS ADJUSTMENT

(RESPONSE RATES: 50% & 70%)

Variable	Domain	"True" Total	$\hat{F}_2$
Births	1	1,240,370	1,242,721
	2	1,903,959	1,896,650
	Aggregate	3,144,329	3,139,371
Divorces	1	373,636	375,951
	2	662,022	661,383
	Aggregate	1,035,658	1,037,334
U.S. Population	1	80,302,039	80,703,819
	2	132,763,646	132,222,890
	Aggregate	213,065,685	212,926,708
Food Stores	1	125,547	121,968
	2	148,355	146,973
	Aggregate	273,902	268,941
Gasoline Service Stations	1	113,772	113,620
	2	118,573	118,541
	Aggregate	232,345	232,161
Hospitals	1	3998	4,117
	2	3326	3,308
	Aggregate	7324	7,426
Marriages	1	942,908	921,443
	2	1,208,807	1,174,342
	Aggregate	2,151,715	2,095,786
Public School Enrollment	1	17,757,679	17,829,987
	2	27,031,137	26,872,449
	Aggregate	44,788,816	44,702,437
Vehicle Thefts	1	127,408	126,140
	2	887,005	872,901
	Aggregate	1,014,413	999,040

TABLE 7.6. COMPARISON OF THE ESTIMATED TOTALS TO THE "TRUE" TOTALS BASED ON EXTERNAL INFORMATION USING POPULATION WEIGHTING CLASS ADJUSTMENT

(RESPONSE RATES: 70% & 90%)

Variable	Domain	"True" Total	$\hat{F}_2$
Births	1	1,240,370	1,229,259
	2	1,903,959	1,905,628
	Aggregate	3,144,329	3,134,887
Divorces	1	373,636	371,185
	2	662,022	661,847
	Aggregate	1,035,658	1,033,033
U.S. Population	1	80,302,039	80,111,780
	2	132,763,646	133,151,385
	Aggregate	213,065,685	213,263,165
Food Stores	1	125,547	122,723
	2	148,355	148,081
	Aggregate	273,902	270,804
Gasoline Service Stations	1	113,772	112,602
	2	118,573	119,237
	Aggregate	232,345	231,840
Hospitals	1	3998	4,012
	2	3326	3,331
	Aggregate	7324	7,343
Marriages	1	942,908	904,138
	2	1,208,807	1,192,378
	Aggregate	2,151,715	2,095,516
Public School Enrollment	1	17,757,679	17,673,031
	2	27,031,137	27,086,337
	Aggregate	44,788,816	44,759,368
Vehicle Thefts	1	127,408	124.647
	2	887,005	876,273
	Aggregate	1,014,413	1,000,920

TABLE 7.7. COMPARISON OF THE ESTIMATED TOTALS TO THE "TRUE" TOTALS BASED ON EXTERNAL INFORMATION USING POPULATION WEIGHTING CLASS ADJUSTMENT

(RESPONSE RATES: 90% & 95%)

Variable	Domain	"True" Total	$\hat{F}_2$
Births	1	1,240,370	1,247,386
	2	1,903,959	1,892,387
	Aggregate	3,144,329	3,139,773
Divorces	1	373,636	375,268
	2	662,022	658,947
	Aggregate	1,035,658	1,034,215
U.S. Population	1	80,302,039	80,855,390
	2	132,763,646	132,298,024
	Aggregate	213,065,685	213,153,414
Food Stores	1	125,547	123,061
	2	148,355	146,848
	Aggregate	273,902	269,909
Gasoline Service Stations	1	113,772	113,027
	2	118,573	118,525
	Aggregate	232,345	231,551
Hospitals	1	3998	4,075
	2	3326	3,305
	Aggregate	7324	7,380
Marriages	1	942,908	907,284
	2	1,208,807	1,171,806
	Aggregate	2,151,715	2,079,090
Public School Enrollment	1	17,757,679	17,870,659
	2	27,031,137	26,917,984
	Aggregate	44,788,816	44,788,643
Vehicle Thefts	1	127,408	127,096
	2	887,005	865,821
	Aggregate	1,014,413	992,917

TABLE 7.8. COMPARISON OF THE ESTIMATED TOTALS TO THE "TRUE" TOTALS BASED ON EXTERNAL INFORMATION USING POPULATION WEIGHTING CLASS ADJUSTMENT

(RESPONSE RATES: 100% & 100%)

Variable	Domain	"True" Total	$\hat{F}_2$
Births	1	1,240,370	1,242,581
	2	1,903,959	1,899,163
	Aggregate	3,144,329	3,141,743
Divorces	1	373,636	374,101
	2	662,022	651,757
	Aggregate	1,035,658	1,035,858
U.S. Population	1	80,302,039	80,504,697
	2	132,763,646	132,691,465
	Aggregate	213,065,685	213,196,162
Food Stores	1	125,547	123,184
	2	148,355	147,512
	Aggregate	273,902	270,696
Gasoline Service Stations	1	113,772	112,867
	2	118,573	118,912
	Aggregate	232,345	231,780
Hospitals	1	3998	4,073
	2	3326	3,317
	Aggregate	7324	7,390
Marriages	1	942,908	903,162
	2	1,208,807	1,185,287
	Aggregate	2,151,715	2,088,449
Public School Enrollment	1	17,757,679	17,800,101
	2	27,031,137	26,996,302
	Aggregate	44,788,816	44,796,403
Vehicle Thefts	1	127,408	126,338
	2	887,005	871,318
	Aggregate	1,014,413	997,657

TABLE 7.9. DISTRIBUTION-FREE SIGN TEST BASED ON EXTERNAL INFORMATION USING SUM OF WEIGHTS ADJUSTMENT

$\hat{V}_3$ vs $\hat{V}_4$

Ho: $\theta = 0$

Ha: $\theta > 0$

Variable	Domain	RESPONSE RATE 50% & 70%	RESPONSE RATE 70% & 90%	RESPONSE RATE 90% & 95%
Births	1			
	2	**	**	**
	Aggregate	**	**	**
Divorces	1			
	2	**	**	**
	Aggregate	**	**	**
U.S. Population	1			
	2	**	**	**
	Aggregate	**	**	**
Food Stores	1			
	2	**	**	**
	Aggregate	**	**	**
Gasoline Service Stations	1			
	2	**	**	**
	Aggregate	**	**	**
Hospitals	1			
	2	**	**	**
	Aggregate	**	**	**
Marriages	1			
	2	**	**	**
	Aggregate	**	**	**
Public School Enrollment	1			
	2	**	**	**
	Aggregate	**	**	**
Vehicle Thefts	1			
	2	**	**	**
	Aggregate	**	**	**

where: θ = median of the differences (Z_{ijk}'s)

$Z_{ijk} = \hat{V}_{3ijk} - \hat{V}_{4ijk}$, i^{th} sample, j^{th} variable, k^{th} domain

*Significant at the .05 level

**Significant at $\leq$ .01 level

TABLE 7.10. DISTRIBUTION-FREE SIGN TEST
BASED ON EXTERNAL INFORMATION
USING POPULATION WEIGHTING CLASS ADJUSTMENT

$\hat{V}_4$ vs $\hat{V}_3$

Ho: $\theta = 0$

Ha: $\theta > 0$

Variable	Domain	Response Rate 50% & 70%	70% & 90%	90% & 95%
Births	1	**	**	**
	2	**	**	**
	Aggregate	**	**	**
Divorces	1	**	**	**
	2	**	**	**
	Aggregate	**	**	**
U.S. Population	1	**	**	**
	2	**	**	**
	Aggregate	**	**	**
Food Stores	1	**	**	**
	2	**	**	**
	Aggregate	**	**	**
Gasoline Service	1	**	**	**
Stations	2	**	**	**
	Aggregate	**	**	**
Hospitals	1	**	**	**
	2	**	**	**
	Aggregate	**	**	**
Marriages	1	**	**	**
	2	**	**	**
	Aggregate	**	**	**
Public School	1	**	**	**
Enrollment	2	**	**	**
	Aggregate	**	**	**
Vehicle Thefts	1	**	**	**
	2	**	**	**
	Aggregate	**	**	**

where: θ = median of the differences (Z_{ijk}'s)

$Z_{ijk} = \hat{V}_{4ijk} - \hat{V}_{3ijk}$, i^{th} sample, j^{th} variable, k^{th} domain

Under Ho: $P(Z_{ijk} > 0) = P(Z_{ijk} < 0) = \frac{1}{2}$

*Significant at the .05 level
**Significant at $\leq$ .01 level

8. SUMMARY

Classical estimators introduced under conditions 1 and 4 are for the most part highly unreliable. These estimators are very sensitive to variation among data. Depending on which of the two types of adjustments is used, the researcher may expect an underestimate or an overestimate. As shown, the response rate and data variation are the most important contributors to the quality of the estimates. Table 8.1 contains some circumstances which have proven to greatly affect the estimate's accuracy; also the resulting quality of the estimates are included. It has been shown that data vary in domain #1 less than in domain #2. Therefore, the degree of variation in data will be determined by the average, by domain, of the percent of the variance estimates to the unbiased estimates.

As can be seen from circumstances I - VIII, $\hat{V}_1$ and $\hat{V}_4$ are less efficient estimators and are very unstable. $\hat{V}_2$ and $\hat{V}_3$ are far more favorable under the above circumstances. This does not by any means imply that these estimators are the best that do exist. However, from the current study it is evident that $\hat{V}_2$ and $\hat{V}_3$ have their merits over $\hat{V}_1$ and $\hat{V}_4$. When weighting class adjustments are used to adjust for nonresponse and the variance is estimated based on an approximation of Taylor's series, it is recommended that $\hat{V}_2$ and $\hat{V}_3$ be used. Therefore, statistical software should be modified to generate these estimates which are more favorable.

TABLE 8.1. CONDITIONAL SUMMARY OF ALL FOUR VARIANCE ESTIMATORS

Circumstance		$\hat{V}_1$	$\hat{V}_2$	$\hat{V}_3$	$\hat{V}_4$
I.					
i)	sum of weights adjustment				
ii)	low response rate	1	0	-2	2
iii)	"little" variation in data				
II.					
i)	sum of weights adjustment				
ii)	low response rate	-3	-2	-1	-3
iii)	"considerable" variation in data				
III.					
i)	sum of weights adjustment				
ii)	high response rate	0	0	-2	1
iii)	"little" variation in data				
IV.					
i)	sum of weights adjustment				
ii)	high response rate	-2	-1	-2	-3
iii)	"considerable" variation in data				
V.					
i)	population weighting class adjustment				
ii)	low response rate	2	-1	-1	3
iii)	"little" variation in data				
VI.					
i)	population weighting class adjustment				
ii)	low response rate	2	-1	-1	3
iii)	"considerable" variation in data				
VII.					
i)	population weighting class adjustment				
ii)	high response rate	0	0	0	3
iii)	"little" variation in data				
VIII.					
i)	population weighting class adjustment				
ii)	high response rate	0	0	0	3
iii)	"considerable" variation in data				

-3 Vastly underestimate (<50%)
-2 Underestimate (50% to <85%)
-1 Slightly underestimate (85% to <95%)
0 Good estimate (95% to <105%)
1 Slightly overestimate (105% to <115%)
2 Overestimate (115% to <150%)
3 Vastly Overestimate (>150%)

LIST OF REFERENCES

1. Barr, A. J. and J. H. Goodnight. 1979. SAS User's Guide, 1979 Edition. SAS Institute, Inc., Cary, North Carolina.

2. Chapman, D. W. 1976. A Survey of Nonresponse Imputation Procedures. Proceedings of the Social Statistics Section. J. American Stat. Association. 1:245-251.

3. Cochran, W. G. 1977. Sampling Techniques, 3rd Edition. John Wiley & Sons, Inc., New York.

4. Department of Economic and Social Affairs Statistical Office of the United Nations. 1964. Recommendations for the Preparation of Sample Survey Reports. Stat. Paper Series C (1), Rev. 2, United Nations, New York.

5. Hansen, M. H., W. N. Hurwitz and W. G. Madow. 1953. Sample Survey Methods and Theory. Methods and Applications. 1:560. John Wiley & Sons, Inc., New York.

6. Hartley, H. O. 1959. Analytic Studies of Survey Data. Instituto di Statistica, Rome, volume in honor of Corrado Gini.

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

8. Kendall, M. G. and A. Stuart. 1968. The Advanced Theory of Statistics, 2nd Edition. Hafner Pub. Co., New York.

9. Kish, L. 1965. Survey Sampling. John Wiley & Sons, Inc., New York. pp. 527-528.

10. National Center for Health Statistics. 1975. Distribution and Properties of Variance Estimators for Complex Multistage Probability Samples - An Empirical Distribution. Vital and Health Statistics. Series 2, No. 65. DHEW Pub. No. (HRA) 75-1339. Health Resources Administraton. Washington, DC. U.S. Government Printing Office.

11. Raj, D. 1964. The Use of Systematic Sampling with Probability Proportional to Size in a Large-Scale Survey. J. American Stat. Assoc. 59:251-255.

12. Shah, B. V. 1979. SESUDAAN: Standard Errors Program for Computing of Standardized Rates from Sample Survey Data. Research Triangle Institute, Research Triangle Park, North Carolina. Prepared for the University of North Carolina, Chapel Hill, North Carolina.

13. Singh, M. P. 1978. Alternative Estimators in PPS Sampling. Survey Methodology. 4(1):264-276.

14. Singh, M. P. and R. Tessier. 1975. Some Estimates for Domain Totals. Survey Methodology. 1(1):74-86.

15. Steel, R. G. D. and J. H. Torrie. 1980. Principles and Procedures of Statistics, A Biometrical Approach, 2nd Edition. McGraw-Hill, Inc., New York. p. 102.

16. Stuart, A. 1962. Basic Ideas of Scientific Sampling. No of Griffin's Statistical Monographs and Courses. Hafner Pub. Co., New York. pp. 92-97.

17. U.S. Bureau of the Census. 1978. The Current Population Survey, A Report on Methodology. Technical Paper No. 40. U.S. Government Printing Office, Washington, DC. pp. 56-58.

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

19. U.S. Department of Commerce. 1976. County and City Data Book 1975. A Statistical Abstract Supplement. U.S. Government Printing Office, Washington, DC.

20. Vaivanijkul, N. 1961. A Comparison of Respondents and Nonrespondents in a Sample Survey. Unpublished M. S. Thesis, North Carolina State University Library, Raleigh, North Carolina.

21. Woodruff, R. S. 1971. A Simple Method for Approximating the Variance of a Complicated Estimate. J. American Stat. Assoc. 66(334):411-414.