IMPROVED VARIANCE ESTIMATORS USING WEIGHTING CLASS ADJUSTMENTS FOR SAMPLE SURVEY NONRESPONSE

Shelton M. Jones and James R. Chromy, Research Triangle Institute

1. INTRODUCTION

This paper is primarily based on work presented in the first author's thesis under the direction of the second author.

Common weighting class 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 for the total number of eligibles in weighting class c, t_{1c},

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

The weighting class adjustment is simply t_{1c}/t_{2c}. Quite often, classical procedures treat both components 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 I: When t_{1c} and t_{2c} are generated from sample information but are treated as constants.

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

Condition III: 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 IV: 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 conventional variance estimates that are based on conditions I and IV. There is no discussion of conditions II and III in previous readings, therefore variance estimators due to Conditions II and III will be referred to as the Jones-Chromy variance estimators. It shall be shown that the Jones-Chromy estimators yield estimates that have better properties and are far more reliable than conventional ones.

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. An example would be to examine a simple random sample (SRS), without replacement 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 CPS 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.

Let $\hat{F}_2$ be a total estimator for the variable of interest Y_3 observed for current survey respondents. Then it follows for any design that:

$$\hat{F}_2 = \sum_{c=1}^{C} \frac{T_{1c}}{t_{2c}} t_{3c} \qquad (2.1)$$

$$\text{where for SRS: } t_{3c} = \sum_{k=1}^{n} \frac{N}{n} Y_{3ck} \qquad (2.2)$$

$$Y_{3ck} = \begin{cases} \text{the value for the variable of interest given sample unit k is a respondent in weighting class c,} \\ 0, \text{ otherwise.} \end{cases}$$

2.1 ESTIMATORS BASED ON SAMPLE INFORMATION

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) adjustment methodology divides 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:

$$\frac{t_{1c}}{t_{2c}} = \frac{\text{(interviewed households + noninterviewed households)}}{\text{interviewed households}} \qquad (2.3).$$

The ratios are uniformly applied to the respondents within each cell as long as the ratios are not greater than or equal to 2.0. If a ratio exceeds this criterion then counts are grouped for all races within the residence category in the cluster.

Common statistical packages such as SESUDAAN[1/] and SUPER CARP[2/], due to Shah (1979) and Hidiroglou (1980), respectively, treat both t_{1c} and t_{2c} as constants in the computation of variance estimates associated with equation 2.1 and,

$$\hat{F}_1 = \sum_{c=1}^{C} \frac{t_{1c}}{t_{2c}} t_{3c} \ . \qquad (2.4)$$

2.2 TAYLOR SERIES VARIANCE ESTIMATION

The delta method or Taylor series linearization variance approximation is widely used among various research agencies to approximate the variance of nonlinear survey statistics. Common goals are to estimate the variance associated with a nonlinear estimator by reducing it to a linear form. The linearization is obtained as a function of first order partial derivatives ignoring second and higher order terms. 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 implemented by the U.S. Bureau of the Census (1978) 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."

2.3 SAMPLING UNITS SELECTED PPS WITH REPLACEMENT

Previous investigations have been largely based on stratified simple random sampling without replacement. It is more often the case that large-scale sample surveys consist of multi-stage designs where the first stage units are selected 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 when the first stage size measures are exact and an equal size second stage sample is drawn from each primary unit.

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 the variance estimator is not complicated by the number of subsequent sampling stages as long as primary units are sampled PPS with replacement and the second and subsequent stage samples are drawn independently within each PSU. To the extent that the size measure is correlated with the characteristic variable of interest, 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 with replacement is that the form of the estimated variance is quite simple, as will be shown later. A further advantage is the variance estimator appropriate for with replacement sampling is a reasonably good approximation when first stage units are selected without replacement. This result was substantiated by Raj (1964). He determined that, if the PPS with replacement variance estimator 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.

An unbiased sample variance estimator for units selected with unequal probabilities with replacement may be obtained from Cochran (1977).

3. THEORETICAL FORMULATION

Estimators for the variance of $\hat{F}_1$ and $\hat{F}_2$ in equations 2.4 and 2.1 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.

As mentioned, an expansion of the Taylor Series will be used to reduce these nonlinear forms to linear forms. As usual, the samples are assumed to be large enough so that remaining terms can be omitted after the first-order approximations. According to Woodruff (1971), if the partial derivatives are evaluated at their expected values then the large sample approximation for the variance of $\hat{F}$ is :

$$\mathrm{Var}\ (\hat{F}) \doteq E\left\{ \sum_{c=1}^{C} \sum_{a=1}^{3} \frac{\partial \hat{F}}{\partial t_{ac}} (t_{ac} - E\ t_{ac}) \right\}^2 . \qquad (3.1)$$

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 can be shown that when all covariances are zero,

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

Efforts will now be made towards deriving an estimator for the variance of $\hat{F}_1$ and $\hat{F}_2$ based on the four conditions which were described and on the sample design. For equaton 2.4,

$$t_{ac} = \sum_{h=1}^{R} \sum_{k=1}^{n_h} t_{achk}, \quad a\varepsilon\{1,2,3\},$$
in equation 2.1, $a\varepsilon\{2,3\}$

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

W_{hki} = the sampling weight corresponding to second stage unit i (SSU-i) within PSU k within stratum h

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

$$Y_{1chki} = \begin{cases} 1, \text{ if SSU i within PSU k within stratum h is in weighting class c} \\ 0, \text{ otherwise.} \end{cases}$$

$$Y_{2chki} = \begin{cases} 1, \text{ if SSU i within PSU k within stratum h is in weighting class c and is a respondent} \\ 0, \text{ otherwise.} \end{cases}$$

$$Y_{3chki} = \begin{cases} \text{the value for the variables of interest given SSU i within PSU k within stratum h is in weighting class c and is a respondent} \\ 0, \text{ otherwise.} \end{cases}$$

An expansion of equation 3.1 is often tedious to work with in practice. A better procedure was demonstrated by Woodruff. Thus, pertaining to Woodruff's procedure, equation 3.1 is equal to:

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

where $T_{ach} = n_h \, E \{ t_{achk} \}$ for all PSUs $k = 1, \ldots, n_h$.

According to Woodruff if the order of summation is reversed the variance can be easily evaluated. Using this method equation 3.3 is equivalent to

$$\mathrm{Var} \left\{ \sum_{h=1}^{R} \sum_{k=1}^{n_h} \sum_{c=1}^{C} \sum_{a=1}^{3} \frac{\partial \hat{F}}{\partial t_{ac}} t_{achk} \right\} \tag{3.4}$$

where the partial derivatives $(\frac{\partial \hat{F}}{\partial t_{ac}})$ are evaluated at estimated values. At this point another variable will be defined; namely,

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

Therefore, equation 3.4 is estimated by,

$$\mathrm{var} \sum_{h=1}^{R} \frac{1}{n_h} \sum_{k=1}^{n_h} \hat{U}_{hk} = \sum_{h=1}^{R} \sum_{k=1}^{n_h} \frac{(\hat{U}_{hk} - \bar{\hat{U}}_h)^2}{n_h(n_h-1)} = \hat{V}_i \tag{3.5}$$

where: $\hat{V}_i$ = the variance estimator derived from the i-th Condition.

$\bar{U}_h$ = the average of $\hat{U}_{hk}$ over the n_h PSUs in stratum h.

Without regard to the sample design, each lower case t_{ac} will denote an estimator for the characteristic of interest. The upper case versions, T_{ac}, will denote known constants.

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:

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

Condition II:

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

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

$$+ n_h \sum_{c=1}^{C} \frac{t_{1c}}{t_{2c}} t_{3chk} \tag{3.7}$$

Condition III:

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

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

Condition IV:

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

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

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 non-responding 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).

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

As described by Jones (1981), two types of weighting class adjustments will be studied:
(1) Sum of weights adjustment and
(2) Sum of weighted population adjustment.

A completely random procedure was used to determine which second stage units would be respondents and which would be nonrespondents. Then the following response rates will apply:

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

The two weighting classes are based on the percent of the county's population that lies in an urbanized area. Counties whose population is 50% urban or less are included in weighting class #1 and greater than 50% in weighting class #2.

Domain #1 consist of those counties whose 1970 population as determined by the 1970 census, was less than 135,000. Domain #2 includes counties whose 1970 population was 135,000 or greater.

The following is a sampling frame 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. RESULTS

Figures 1-4 contain plots of the average precent relative standard error (RSE) versus the weighted average response rate. The RSE's are given only for the aggregates and are computed as a function of the "true total" since the actual total is known. Given T = the "true total" based on the sampling frame, it follows that:

$$RSE_{(1)} = \frac{\hat{V}_{(1)}^{\frac{1}{2}}}{T} \cdot 100,$$ if the variance of $\hat{F}_1$ (i.e., $\hat{V}_{(1)}$) is estimated by replicated samples

$$RSE_{(2)} = \frac{\hat{V}_{(2)}^{\frac{1}{2}}}{T} \cdot 100,$$ if the variance of $\hat{F}_2$ (i.e., $\hat{V}_{(2)}$) is estimated by replicated samples

$$RSE_i = \frac{\bar{V}_i^{\frac{1}{2}}}{T} \cdot 100,$$ where i = 1,...,4, which corresponds to the i-th Condition previously mentioned

and $\bar{V}_i = \frac{\sum_{j=1}^{100} \hat{V}_{ij}}{100}$, where $\hat{V}_{ij}$ is the variance estimate for the i-th Condition based on the j-th simulation.

Once the RSE's are computed, averaging is done over all 9 variables for the aggregates.

Described in Figures 5-8 are plots of the average relative bias versus the weighted average response rate for the aggregate. The relative biases are derived by subtracting the corresponding replicated variance estimate from the i-th variance estimate and dividing by the replicated estimate.

Estimators $\hat{V}_{(2)}$, $\bar{V}_3$, $\bar{V}_4$ utilize T_{1c} which is based on an external source and is assumed to be free of random error. It is then feasible to compare these estimators on the same plot. Theoretically, $\hat{V}_{(2)}$, is both unbiased and consistent, therefore it's estimates are used as the standard for measuring the accuracy of $\bar{V}_3$ and $\bar{V}_4$. For the plots in Figure 1 and 3, the distribution of average percent RSE's of $\bar{V}_3$ are more similar to those of $\hat{V}_{(2)}$ as compared to $\bar{V}_4$ versus $\hat{V}_{(2)}$. This is true through 100% response. As is expected the average percent RSE decreases as the response rate increases. In terms of relative bias, as indicated in Figures 5 and 7, $\bar{V}_3$ is less bias than $\bar{V}_4$. As seen in Figure 5 for the low response rate, $\bar{V}_4$ has a positive bias of over 600% as compared to $\bar{V}_3$'s small negative bias of less than 10%.

A similar comparison is made for $\hat{V}_{(1)}$, $\bar{V}_1$, $\bar{V}_2$ since these variance estimators are to some extent functions of t_{1c}. As it was for post-stratification, according to Figure 6, a positive bias of over 600% exists for the low response rate when $\bar{V}_1$ is used. This compares to less than a 10% negative bias obtained from $\bar{V}_2$.

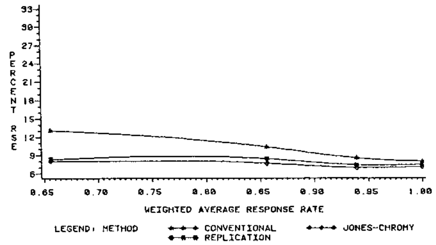
6. CONCLUSIONS

Findings have revealed that variance estimates obtained from the conventional methods (Conditions I and IV) were for the most part significantly biased. These estimators proved to react in accordance to amount of variation among data and the severity of nonresponse. Also, the weighting class specifications had the single most important effect on the estimators' accuracy. The Jones-Chromy variance estimators derived from Conditions II and III are less affected by these circumstances and are far more stable.

These conclusions hold true and are substantiated in Jones' previous work. It is further emphasized that when weighting class adjustments are used to adjust for nonresponse and the variance is estimated using Taylor's series variance approximation, that the Jones-Chromy variance estimators should be utilized. Therefore, computer algorithms should be constructed and statistical software modified to incorporate these two estimators.

Figure 1.

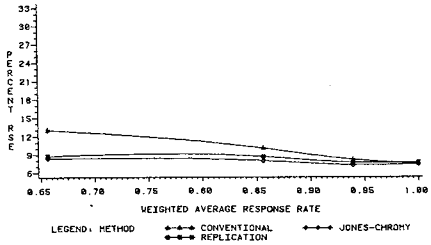
AVERAGE PERCENT RSE
USING POPULATION ADJUSTMENT
BASED ON POST-STRATIFICATION



Variance Estimators	Response Rates			
	Low	Medium	High	100%
Replication - $\hat{V}_{(2)}$	8.4	8.4	7.4	7.5
Jones-Chromy - $\tilde{V}_3$	8.0	7.6	6.9	7.1
Conventional - $\bar{V}_4$	13.1	10.3	8.5	7.9

Figure 2.

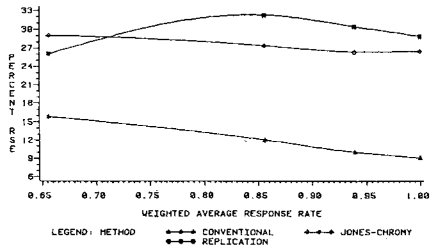
AVERAGE PERCENT RSE
USING POPULATION ADJUSTMENT
BASED ON SAMPLE RESULTS



Variance Estimators	Response Rates			
	Low	Medium	High	100%
Replication - $\hat{V}_{(1)}$	8.8	8.6	7.6	7.7
Jones-Chromy - $\tilde{V}_2$	8.4	7.9	7.2	7.4
Conventional - $\bar{V}_1$	13.0	10.0	8.1	7.4

Figure 3.

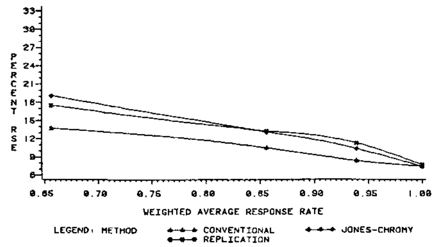
AVERAGE PERCENT RSE
USING SUM OF WEIGHTS ADJUSTMENT
BASED ON POST-STRATIFICATION



Variance Estimators	Response Rates			
	Low	Medium	High	100%
Replication - $\hat{V}_{(2)}$	26.1	32.2	30.3	28.8
Jones-Chromy - $\tilde{V}_3$	29.0	27.3	26.2	26.4
Conventional - $\bar{V}_4$	15.8	11.9	9.9	9.0

Figure 4.

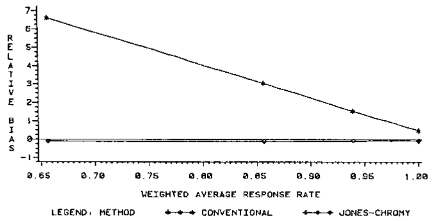
AVERAGE PERCENT RSE
USING SUM OF WEIGHTS ADJUSTMENT
BASED ON SAMPLE RESULTS



Variance Estimators	Response Rates			
	Low	Medium	High	100%
Replication - $\hat{V}_{(1)}$	17.5	13.2	11.2	7.7
Jones-Chromy - $\tilde{V}_2$	19.1	13.0	10.3	7.4
Conventional - $\bar{V}_1$	13.7	10.4	8.3	7.4

Figure 5.

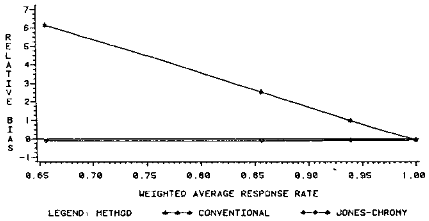
AVERAGE RELATIVE BIAS
USING POPULATION ADJUSTMENT
BASED ON POST-STRATIFICATION



Variance Estimators	Response Rates			
	Low	Medium	High	100%
Jones-Chromy - $\tilde{V}_3$	-0.096	-0.111	-0.078	-0.062
Conventional - $\bar{V}_4$	6.596	3.047	1.543	0.472

Figure 6.

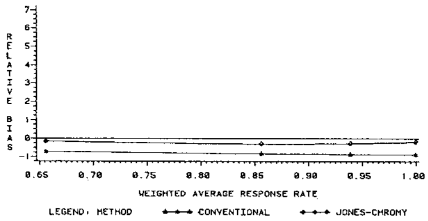
AVERAGE RELATIVE BIAS
USING POPULATION ADJUSTMENT
BASED ON SAMPLE RESULTS



Variance Estimators	Response Rates			
	Low	Medium	High	100%
Jones-Chromy - $\tilde{V}_2$	-0.096	-0.102	-0.064	-0.048
Conventional - $\bar{V}_1$	6.143	2.551	0.977	-0.048

Figure 7.

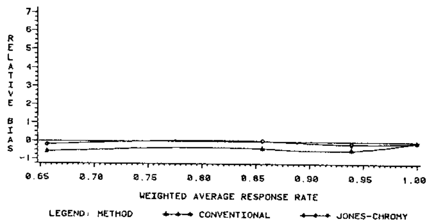
AVERAGE RELATIVE BIAS
USING SUM OF WEIGHTS ADJUSTMENT
BASED ON POST-STRATIFICATION



Variance Estimators	Response Rates			
	Low	Medium	High	100%
Jones-Chromy - $\tilde{V}_3$	-0.145	-0.283	-0.255	-0.163
Conventional - $\bar{V}_4$	-0.716	-0.842	-0.873	-0.875

Figure 8.

AVERAGE RELATIVE BIAS
USING SUM OF WEIGHTS ADJUSTMENT
BASED ON SAMPLE RESULTS



Variance Estimators	Response Rates			
	Low	Medium	High	100%
Jones-Chromy - $\tilde{V}_2$	-0.173	0.024	-0.142	-0.048
Conventional - $\bar{V}_1$	-0.569	-0.400	-0.479	-0.048

FOOTNOTES

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

[2] SUPER CARP is a software package that computes regression equations with known measurement error variance and regression equation with known reliability. Common survey estimators and their estimated variances are also computed.

REFERENCES

1. Chapman, D. W. 1976. A Survey of Nonresponse Imputation Procedures. Proceedings of the Social Statistics Section. J. American Stat. Association. 1:245-251.
2. Cochran, W. G. 1977. Sampling Techniques, 3rd Edition. John Wiley & Sons, Inc., New York.
3. 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.
4. Hidiroglou, M. A., W. A. Fuller and R. D. Hickman. 1980. SUPER CARP, 6th Edition. Survey Section, Statistical Laboratory, Iowa State University, Ames, Iowa.
5. Jones, S. M. 1981. Improved Variance Estimators Using Weighting Class Adjustments for Sample Survey Nonresponse. M.S. Thesis, North Carolina State University Library, Raleigh, North Carolina.
6. 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 Administration. Washington, D.C. U.S. Government Printing Office.
7. 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.
8. 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.
9. Singh, M. P. 1978. Alternative Estimators in PPS Sampling. Survey Methodolgy. 4(1): 264-276.
10. Stuart, A. 1962. Basic Ideas of Scientific Sampling. No 5 of Griffin's Statistical Monographs and Courses. Hafner Pub. Co., New York. pp. 92-97.
11. 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, D.C. pp. 56-58, 153.
12. Woodruff, R. S. 1971. A Simple Method for Approximating the Variance of a Complicated Estimate. J. American Stat. Assoc. 66(334): 441-414.