A COMPARISON OF ALTERNATIVE VARIANCE ESTIMATION STRATEGIES FOR COMPLEX SURVEY DATA

Vicki L. Burt and Steven B. Cohen, NCHSR

Introduction

Data from complex survey designs require special consideration with regard to variance estimation and analysis, due to the violation of simple random sampling assumptions. This is a consequence of design components which include unequal selection probabilities, stratification and clustering. Specially designed software packages exist, which allow for the generation of appropriate variance estimates for statistics derived from complex survey data. These statistics are most often expressed in terms of means, totals, ratios, proportions and regression coefficients. The variance estimation methods used by these statistical packages include balanced repeated replication, jack-knife and Taylor series linearization (Kish and Frankel, 1974). When many tables are required, the computation cost can be large if these procedures are used for each estimate of interest. In addition, the inclusion of variance estimates for all statistics in a data report would often yield cumbersome and expensive documents.

This paper focuses on alternative methods of variance estimation, which consider the complexities of the survey design in a cost efficient manner. Attention is given to three techniques which are respectively referred to as the relative variance curve strategy, the average relative standard error model, and the average design effect model. These methods depend on direct variance estimates for only a representative subset of parameter estimates under consideration. One can then determine variance estimates for all related statistics by applying the respective prediction strategy.

The model specifications for the alternative variance estimation techniques are presented in detail. In addition, the accuracy of these alternative strategies are compared for a representative set of survey statistics specific to data from the National Medical Care Expenditure Survey (NMCES), which has a complex design. Further, the potential of each technique as a method of summarization for purposes of publication is compared.

Relative Variance Curve Strategy

Variance estimates were not computed for each statistic considered in the NMCES by direct methods, due to the constraints of computation time and cost. Another consideration was that inclusion of all relevant variance estimates in NMCES data reports would yield cumbersome documents. The relative variance curve strategy was considered as one alternative approach. It depends on variance estimates for only a representative subset of all parameter estimates under consideration, which are derived by one of the direct methods appropriate for complex survey data. The subset of statistics should be characterized by domains whose underlying demographic characteristics insure a wide range of variability in the parameter estimates.

The relative variance curve strategy for aggregate statistics considers the empirically determined inverse relationship between the size of an estimate Y and its relative variance. This relationship (Bean, 1970) is expressed as:

$$\text{Rel var }(\hat{Y}) = \frac{\sigma^2_{(\hat{Y})}}{\hat{Y}^2} \doteq \alpha + \frac{\beta}{\hat{Y}}$$

and is estimated as

$$\text{Rel var }(\hat{Y}) = \frac{\sigma^2_{(\hat{Y})}}{\hat{Y}^2} \doteq a + \frac{b}{\hat{Y}}$$

where the regression estimates a and b can be determined by an iterative or weighted least squares procedure (Cohen 1979). A relative standard error curve can then be derived by taking the square root of the relative variance curve

$$\text{Rel SE }(\hat{Y}) \doteq \sqrt{a + \frac{b}{\hat{Y}}}$$

To illustrate the method, a curve was fitted to estimates of the number of insured persons from different demographic subgroups of the U.S. population, and their relative variances. The respective relative standard error curve is shown in illustration 1 (not presented due to space limitations) where estimated values of α and β are a=.001732 and b=32,311.84.

This relationship is also used to derive relative variance estimates of percentages, where the numerator is a subclass of the denominator. Here it can be shown that the relative standard error of a percent estimate p, where

$$\hat{p} = \frac{\hat{Y}}{\hat{T}} 100$$

takes the form

$$\text{Rel SE }(\hat{p}) \doteq \sqrt{\frac{b}{\hat{T}} \frac{(100-\hat{p})}{p}}$$

where b is the estimated coefficient determined in the curve fitting procedure for aggregate statistics, and T is the estimated population base. Consequently, the variability of percentage estimates depends on both the respective population base and the percent value. Illustration 2 (not presented due to space limitations) presents the respective relative standard error curves for percentages, again considering characteristics of the insured population. Linear interpolation is used when the population base of the percent is between values specified for the relative standard error curves.

Variances of ratio estimators are derived by considering the relationship that specifies the relative variance of a ratio as approximately equivalent to the sum of the relative variances of the numerator and denominator (French, 1978). More specifically, consider the ratio estimator $R = X/Y$ where the numerator is not a subclass of the denominator. This relation takes the form:

$$\text{Rel var }(\hat{R}) \doteq \text{Rel var }(\hat{X}) + \text{Rel var }(\hat{Y})$$
$$\doteq a_1 + \frac{b}{\hat{X}} + a_2 + \frac{c}{\hat{Y}}$$
$$\doteq a + \frac{b}{\hat{R}\hat{Y}} + \frac{c}{\hat{Y}}$$

and the relative standard error is approximated by

$$\text{Rel SE }(\hat{R}) \doteq \sqrt{a + \frac{b}{\hat{R}\hat{Y}} + \frac{c}{\hat{Y}}}$$

Here the variability of the ratio estimator is inversely related to the size of the respective population base and the ratio estimate.

Several alternative curve fitting procedures with different optimization criteria have been considered for estimating model coefficients. These include a weighted least squares estimation strategy, and an iterative procedure that minimizes the relative squared derivations of predicted and observed relative variance estimators (Cohen, 1979).

Once the model coefficients are determined, variances can be predicted for all related statistics through application of a conversion factor to the resultant relative variance estimates. This conversion factor is the estimated statistic squared.

In addition to savings obtained in variance computation costs, the inclusion of summary relative standard error tables in statistical reports as an alternative to publication of a variance estimate for each statistic that is presented, results in marked savings in production costs. To illustrate the presentation of these summary tables, Table 1 contains approximate relative standard errors (expressed as a percent) for estimated population totals.

Example: An estimate of 26.3 million persons in the U.S. population with at least one purchase or repair of eyeglasses or contact lenses has a relative standard error of approximately 3.5 percent (Table 1). The standard error of this estimate then is

$$SE(T) = \frac{26{,}300{,}000(3.5)}{100} = 920{,}500$$

Average Relative Standard Error Model

The relative variance curve strategy was primarily intended to serve as an approximation technique for aggregate statistics expressed in terms of population totals. As noted, it assumes that there is an inverse relationship between the size of an estimate and its relative variance. When the technique is applied to statistics expressed in ratio form, particularly when the numerator is not a subclass of the denominator, the specified procedure will result in an upper bound on the standard error of the statistic and often will overstate the error.

Another strategy presented in this study relies on the functional relationship of the relative standard error of a parameter estimate and the size of the estimate. The method is referred to as the average relative standard error model. Strata are formed based on the size of the parameter estimate for a population domain of interest and the related population estimate. For each domain estimate of an analytical variable of interest, statistic_i, the relative standard error estimated from a direct variance estimation strategy is specified as:

$$RSE_i \doteq SETay_i/\text{statistic}_i$$

where $SETay_i$ is the standard error of statistic_i, approximated by the Taylor series linearization method for our comparisons. The mean relative standard error for each stratum ($MRSE_s$) is then calculated as:

$$MRSE_s = \sum_{i=1}^{m} RSE_i / m$$

where m is the number of domain estimates in stratum s. This mean relative standard error is applied to parameter estimates falling into this stratum, based on the size of the parameter estimate and related population estimate. The standard error, taking the complex sample design into consideration, was approximated by:

$$SEmnrse_i = \text{statistic}_i \times MRSE_s$$

In order to approximate a standard error using this method, a table of mean relative standard errors is required. The respective strata are defined by cross-classification of boundaries defining the parameter estimate and related population total estimate. To illustrate the summarization procedure, the following example is presented using prescribed medicine data (Burt, 1983). The stratification scheme was implemented by constructing quartile boundaries on the size of the domain estimates of the mean number of prescribed medicines, and quintile boundaries on the respective population total estimates. It should be noted that alternative stratification schemes are appropriate. Further, the mean number of prescribed medicines for individuals living alone, with at least one prescribed medicine, and less than 12 years of education is 13.87. The population total estimate for this domain is 5,876,000. The mean relative standard error for estimates of the mean number of prescribed medicines falling in the stratum bounded by population estimates greater than 3,819,000 and less than or equal to 10,656,000, and by parameter estimates greater than 8.971, is .051. (Table 2) Thus, the approximated standard error is:

$$SEmnrse_i = 13.87 \times .051 = .71$$

Average Design Effect Model

The design effect method is another one of a growing set of strategies that have been used as alternatives to the direct methods of variance estimation for complex survey data. The design effect is defined as the ratio of the true variance of a statistic to the variance derived under simple random sampling assumptions (srs). When the design effect is known, variances derived under simple random sampling assumptions can be corrected by a simple multiplication with this factor. For those data bases which originated for complex national survey designs, it is not unusual to experience sample sizes in excess of 10,000 individuals, as observed in the National Medical Care Expenditure Survey. A considerable reduction in computation time and cost is to be expected for these large data bases when variance procedures with srs assumptions are implemented through statistical package programs, as an alternative to the Taylor Series linearization method, balanced repeated replication or the jack-knife method. The only additional requirement is that availability of design effect factors to appropriately adjust for sample design complexities.

Since the design effect is generally a stable measure, with a much narrower range of dispersion than direct variance estimates, use of an average design effect offers a reasonably good trade-off in cost savings for the attendant reduction in accuracy. The accuracy of the average design effect method may be inferred from the level of dispersion characterizing the design effects for a set of related statistics. The introduction of stratification procedures to the average design effect methodology is one potential strategy for achieving gains in precision for the resultant estimates. The strategy has been suggested by Kish and Frankel (1974), as an alternative to the process of directly computing sampling errors for all the different survey variables under investigation and for their respective subclasses. Direct estimates of the design effects are computed for a subset of statistics, and their mean applied to the entire set of related statistics. It is noted that design effects differ for different statistics, for different variables, and for different survey designs. Consequently, the strategy should be separately applied for groups of related variables for the particular survey at hand.

In application, a criterion variable is selected (i.e., medical expenditures) and several domain estimates of this criterion variable are produced. The domain estimates are defined by marginal and cross-classified distributional categories of predetermined demographic meaasures (i.e., mean annual medical expenditures for specific age-race-sex-income classes of the U.S. population). For a representative subset of the domain estimates which characterize the specified criterion variable, direct estimates of the design effects are derived, and an average design effect, is determined in the following manner:

$$\text{Deff} = \sum_{i=1}^{n} \frac{\text{Deff}_i}{n}$$

where Deff_i is a direct estimate of the design effect for domain estimate i, n is the number of domain estimates selected, and Deff is the average design effect. A weighted average may also be considered, where each design effect is weighted by the population estimate for the domain it represents. Here,

$$\text{Deff}_w = \frac{\sum_{i=1}^{n} w_i \, \text{Deff}_i}{\sum_{i=1}^{n} w_i}$$

where w_i is the population total estimate for domain i.

Once the average design effect is determined, variance estimates for all related statistics derived under srs assumptions can be adjusted by this factor, to account for the complexities of a particular survey design. (S^2 complex = Deff • S^2 srs) The cost savings in computer time and dollars are marked, when considering the various permutations of data presentation that are relevant to a diverse user population.

Comparison of the Accuracy Between Methods

To provide for a comprehensive investigation, the accuracy of the alternative methods of variance estimation were compared for a representative set of survey statistics which estimate medical care utilization, and expenditures of the U.S. population using data from the National Medical Care Expenditure Survey. The survey statistics under investigation were all expressed in ratio form. The utilization measures include the number of physician visits, hospital admissions and number of prescribed medicines (Bonham and Corder, 1981). More specifically, physician visits consisted of all ambulatory physician contacts, excluding telephone calls. Hospital admissions included admissions of less than 24 hours and those for women giving birth. Newborns were not counted as separate admissions unless they were admitted separately following delivery. Prescribed medicines included any drug or other medical preparation prescribed by a physician, including refills. Expenditures data for each of these utilization measures were also considered: physician visit expenditures, total expenditures for prescribed medicines, and total expenditures for all hospital admissions (with charges excluded for inpatient physician services). In addition, the domain defining demographic measures for the survey statistics under consideration included age (<5, 5-14,...55-64, 65+), race (white, nonwhite), sex (male, female), health status (excellent, good, fair, poor), marital status (<17, never married, married, widowed, separated, divorced), years of education (0-8, 9-11, 12, 13-15, 16+, under 17 years of age), employment status (worked, unemployed, not in labor force, <16) and size of city (SMSA, non-SMSA).

The diverse set of selected criterion variables also served to represent three distinct classes of survey statistics: narrow, medium and wide range. More specifically, the class of narrow range statistics was determined by data at the individual level, whose measurements generally fall within the range of 0-3. These measurements usually serve to indicate the presence or absence of a population attribute or its frequency of occurrence.

Similarly, medium range statistics consist of measurement which infrequently fall outside the range of 0-10. Wide range statistics are characterized by data more continuous in nature that have much higher upper bounds.

The class of narrow range statistics is represented by NMCES data on the number of hospital admissions. Data on ambulatory visits and number of prescribed medicines served to represent the medium range class. The class of wide range statistics is represented by the following measures: total expenditures for hospital admissions, physician expenditures, and total expenditures for prescribed medicines.

For each of the selected criterion variables, domain estimates were generated in terms of population means or proportions when appropriate. The domain estimates are defined by marginal and cross-classified distributional categories of the selected demographic measures. For example, consider the mean annual expenditures for ambulatory physician visits within specific age-race-sex-health status classes of the U.S. population. The domain estimate, $\overline{Y}_g$, is derived as:

$$\overline{Y}_g = \frac{\sum_{i \epsilon g} W_i X_{gi} Y_i}{\sum_{i \epsilon g} W_i X_{gi}}$$

where Y_i is the ith individual's expenditures for physician visits,

W_i is the ith individual's sampling weight, expressed as the reciprocal of its selection probability and multiplied by nonresponse and post-stratification adjustments, and

X_{gi} = 1 if the individual is a member of the gth age-race-sex-health status domain,

=0 otherwise.

In this study, model parameters for the relative variance curve method were estimated by the weighted least squares optimization strategy. The unweighted estimator was used for the average design effect model. An examination of the design effects for domain estimates of the NMCES health care utilization, and expenditure measures indicated that design effect variation was influenced by sample size and the size of the domain estimate (Cohen, 1983). For purposes of comparison, the same stratification scheme considered for the relative standard error model was adopted. Within each of these twenty distinct strata, the average design effect was computed and used for estimating standard errors. All direct variance estimates for survey statistics were generated through the Taylor Series linearization method. Modelling for each of the specified variance approximation strategies was done separately for the different sets of criterion variables.

To measure the accuracy of the respective variance estimation strategies, the average relative absolute difference between direct and predicted standard error estimates for domain specific population estimates, was considered. The measure took the form:

$$\overline{A} = \sum_{i=1}^{n} \frac{|\hat{S}_{pi} - \hat{S}_{oi}|}{n \hat{S}_{oi}}$$

where S_{oi} is the standard error estimated by the Taylor Series linearization method for the i-th domain specific population estimate, S_{pi} is the standard error predicted by either the average design effect or relative variance curve method for the i-th domain specific population estimate, and n is the number of domain estimates that constitute a representative subset for the criterion variable of interest.

Table 3 presents the comparisons in accuracy for the alternative variance estimation techniques: the relative variance curve strategy, the average relative standard error model, and the average design effect model with stratification. Study findings revealed a consistently lower average relative absolute difference, ($\overline{A}$), for both the average relative standard error model and the design effect model, over the relative variance curve technique. All observed improvements in accuracy were significant at the .05 level as determined by comparisons of confidence intervals. The improvements in accuracy were most prominent for the prescribed medicine and ambulatory visit related parameter estimates. The stratified design effect model performed significantly better than the average relative standard error model, except for estimates of prescribed medicines and their related expenditures.

The order of magnitude observed in the accuracy measure for the relative variance curve strategy was disturbing. The technique has gained a degree of respectability as a consequence of its theoretical justification and widespread usage among a large statistical audience. Given the potential costs incurred by application of one of the direct methods of variance estimation appropriate for complex survey data, some users are willing to accept modest levels of bias that result when alternative cost-effective estimation strategies are applied. The consistent improvement in accuracy obtained by the design effect model and the average relative standard error technique, argues that greater scrutiny must be given to the relative variance curve strategy prior to a decision for adoption.

As noted, the inclusion of all relevant standard error estimates in NMCES data reports, or in data reports summarizing survey findings from designs of similar complexity, would result in cumbersome documents with increased publication costs. Consequently, the potential of each technique as a method of summarization for publication purposes was also compared. The relative variance curve is relatively inexpensive to implement and publish. The technique offers considerable advantages in terms of streamlining publications, by presentation of summary tables as an alternative to point estimates of standard errors. It requires the user to interpolate, and necessitates reference to two summary tables in order to derive standard error approximations for statistics expressed in ratio form, when the numerator is not a sub class of the denominator. Its poor performance in terms of accuracy, further limits its desirability.

The average design effect model with stratification will also yield substantial reductions in computer cost as a variance approximation strategy with superior performance in terms of accuracy. A summary table of average design effects for parameter estimates, stratified by size and population total estimates, is specified. However, the technique requires the availability of the respective standard error estimate under simple random sampling assumptions, except for estimates of percentages. As a result, the techique's potential as summarization procedure is severely limited.

In comparison, the average relative standard error model will yield substantial savings in computer costs as a variance approximation strategy and also in terms of streamlining publications. Only one summary table of average relative standard errors must be referenced for a specified parameter estimate, and the calculation to obtain the standard error approximation is direct, without requiring interpolation. Finally, the technique's relative performance in terms of accuracy does not diminish its desirability.

One additional criterion in evaluating a potential standard error approximation strategy is: does one obtain consistent results for hypothesis tests using standard errors derived from a direct variance estimation strategy. Hypothesis tests using standard errors derived by the Taylor series linearization method were carried out to compare parameter estimates characterizing adults living alone and those living in families for several health care utilization and expenditure measures (Burt, 1983). The same tests for equivalence in estimates were replicated using variances approximated by the average relative standard error model. (Table 4). The vast majority of tests yielded consistent results across variance estimation strategies. Incidences of anti-conservative results obtained by the relative standard error method were rare.

Summary

Several direct methods of variance estimation appropriate for complex survey data have been developed. Use of the direct procedures, which include the Taylor series linearization method, balanced repeated replication, and the jack-knife method, would incur substantial computer costs if applied to each parameter estimate of interest for a large number of tables. In addition, the inclusion of variance estimates for all statistics in a data report would often yield cumbersome documents and increment publication costs. Consequently, alternative variance approximation strategies have been considered as cost effective techniques and for purposes of summarization.

This study has concentrated on three alternative variance approximation strategies that have gained recognition through use: the relative variance curve, the average relative standard error model and the average design effect model. The accuracy of these methods have been compared for a representative set of survey statistics expressed in ratio form, specific to data from the National Medical Care Expenditure Survey, which has a complex design. The potential of each technique as a method of summarization was also compared.

Although the relative variance curve method offers considerable advantages in terms of streamlining publications as a method of summarization for variance approximations, its poor performance in terms of accuracy limits its desirability. The average design effect technique was superior in terms of accuracy but limited in terms of yielding less cumbersome documents. Finally, the average relative standard error technique was the preferred method of summarization and did not depart markedly in terms of accuracy from the performance of the average design effect variance approximation strategy.

The results demonstrate that the decision concerning the variance approximation method for adoption should not be arbitrary. Measures of accuracy should be defined, and the behavior of the alternative techniques under investigation should be compared for a representative subset of survey estimates. In addition, the potential of each method as a method of summarization should be considered. The variance approximation strategy which displays the overall superior performance for the specified criteria should then be selected.

REFERENCES

Bean, J.A.
1970 Estimation and Sampling Variance in the Health Interview Survey. National Center for Health Statistics Vital and Health Statistics, PHS Pub. No. 1000, Series 2 - No. 28. Washington: Government Printing Office.

Burt, V.L.
1983 "Analyses of the Health Care Experience for Single Person Families Using Data from the National Medical Care Expenditure Survey. "Proceedings of the American Statistical Association, Social Statistics Section.

Bonham, G. S. and Corder, L. S.
1981 National Medical Care Expenditure Survey Household Interview Instruments. National Center for Health Services Research, National Medical Care Expenditure Survey, Instruments and Procedures Series, Department of Health and Human

Services Publication No. (PHS) 80-3280, Washington: Government Printing Office.

Cohen, S. B.
1979 "An Assessment of Curve Smoothing Strategies Which Yield Variance Estimates for Complex Survey Data," 1979 Proceedings of the American Statistical Association Survey Research Section, pp. 101-104.

Cohen, S. B. and W. D. Kalsbeek,
1981 National Medical Care Expenditures Survey: Estimation and Sampling Variances in the Household Survey, National Center for Health Services Research, Instruments and Procedures Series, No. 2, Department of Health and Human Services Publication No. (PHS) 81-3281. Washington: Government Printing Office.

Cohen, S. B.
1983 "Design Effect Variation in the National Medical Care Expenditure Survey," Proceedings of the American Statistical Association, Survey Research Section, in press.

French, D. K.
1978 National Survey of Family Growth, Cycle 1. Sample Design, Estimation Procedures, and Variance Estimation. National Center for Health Statistics, Data Evaluation and Methods Research, PHS Pub. No. 78-1350, Series 2, No. 76. Washington: Government Printing Office.

Kish, L. and Frankel, M. R.
1974 "Inferences From Complex Surveys," Journal of the Royal Statistical Society, 36:1-37.

Table 1: Relative Standard Errors for Estimated Population Totals

Estimated population totals (in thousands)	Relative standard error (%)
500	18.9
1,000	13.5
2,500	8.7
5,000	6.4
10,000	4.8
25,000	3.5
50,000	3.0
100,000	2.6
200,000	2.4

SOURCE: National Center for Health Services Research

Table 2. Mean relative standard errors for prescriptions (NMCES, 1977)

Population estimate (Quintile boundaries) (in thousands)	Mean number of prescribed medicines (Quartile boundaries)			
	LE 4.145	GT 4.145 and LE 5.503	GT 5.503 and LE 8.971	GT 8.971
<1,000	—	—	.124	.122
GT 1,000 and LE 3,819	.115	.109	.098	.081
GT 3,819 and LE 10,656	.077	.074	.066	.051
GT 10,656 and LE 27,095	.053	.056	.044	.031
GT 27,095	.032	.032	.030	.023

—Indicates no observations in stratum.

SOURCE: National Center for Health Service Research

Table 3. Average relative absolute differences for variance approximation models.

Class of Statistic	Average relative absolute differences		
Population means	Relative variance curve	Average RSE model	Stratified design effect model
Narrow range			
hospital stays	.080 (.059)	.186 (.014)	.080 (.008)
Medium range			
ambulatory visits	1.827 (.114)	.204 (.016)	.127 (.010)
Prescribed medicines	2.195 (.129)	.176 (.012)	.228 (.012)
Wide range			
Hospital expenditures	0.507 (.053)	.213 (.016)	.083 (.008)
Ambulatory visit expenditures	1.242 (.074)	.219 (.016)	.117 (.010)
Prescription expenditures	0.932 (.061)	.180 (.012)	.224 (.012)

1.96 x standard error in parenthesis
Average calculated over 425 domains.

SOURCE: National Center for Health Services Research

Table 4. Results of Z-tests using standard errors approximated from mean relative standard error model compared to using the Taylor series linearization method.[4]

Class of statistic	Domains	Conservative[1]	Same[2]	Anticonservative[3]
Means per person with event				
Hospital stays	19	0	18	1
Hospital expenditures	19	0	18	1
Prescribed medicines	20	1	19	0
Prescription expenditures	20	1	18	1
Ambulatory visits	21	1	20	0
Ambulatory expenditures	21	4	17	0

[1]Conservative-asymptotic Z-statistic is LT |1.96| using average RSE when it had been GE |1.96| using Taylor series standard error.
[2]The asymptotic Z-statistic is LT |1.96| or GE |1.96| for both methods.
[3]Anticonservative-symptotic Z-statistic is GE |1.96| using average RSE when it had been LT |1.96| using Taylor series standard error.

SOURCE: National Center for Health Services Research