VARIANCE ESTIMATION FOR PUBLIC USE FILES UNDER CONFIDENTIALITY CONSTRAINTS

Wesley Yung, Statistics Canada
R.H. Coats Building, 3-B, Tunney's Pasture, Ottawa, Ontario, K1A 0T6

Key Words: Confidentiality; Public use files; Variance estimation

1. Introduction

Many of Statistics Canada's surveys produce a Public Use Micro-data File (PUMF) which is made available to analysts wishing to perform their own analysis of Statistics Canada data. On these micro-data files, each record represents a sampled element (business, household, etc...) and includes a weight which usually incorporates adjustments for nonresponse and benchmarking. However, as part of disclosure avoidance procedures, design information such as stratum or cluster identifiers are not normally included on the PUMF. In the absence of this design information, users of Statistics Canada's PUMF's are unable to calculate valid design-based variance estimators.

Currently, users are informed of sampling variability by means of 'Approximate Sampling Variability Tables'. These tables give an approximate coefficient of variation (CV) for estimates of totals, ratios and proportions for categorical variables. Unfortunately, these tables can not be used to obtain CV's for continuous variables or for complex statistics such as estimated regression parameters. As well, this approach, in use since the 1970's, is now felt to be unsatisfactory for both ease of use and statistical reasons.

In this paper, the use of the bootstrap method is proposed as a solution to the problem of estimating valid variance estimates from PUMF's while still respecting confidentiality constraints. The bootstrap method can be used to calculate variance estimates for totals, ratios and proportions for categorical and continuous variables, as well as for complex statistics such as regression coefficients and quantiles. In section 2, the construction of the approximate sampling variability tables and consequences of their construction are discussed. Alternative methods investigated at Statistics Canada are presented in section 3 while the proposed bootstrap method is presented in section 4. Comparisons of the proposed method with the approximate sampling variability table method, using data from Statistics Canada's National Population Health Survey, are presented in section 5.

2. Approximate Sampling Variability Tables

Approximate sampling variability tables, or CV look-up tables, have been used for many years as a means of informing micro-data file users of sampling variability. Typically, these tables are produced at both the national and provincial levels and occasionally at sub-provincial levels. For each table, a set of approximately 30 key categorical variables is identified and exact variances are calculated for each response category cross-classified by age groups and sex. As well, variances are calculated under a simple random sample design and design effects (DEFF's) are obtained for each combination of response category, age group and sex. The 75th percentile of these DEFF's is then used as a representative DEFF for use in preparing the CV look-up tables. Use of the 75th percentile means that the estimated CV will be an overestimate 75% of the time and an underestimate the remainder of the time.

Users of Statistics Canada's PUMF's have expressed a wide range of views concerning the use of the CV look-up tables. Unsophisticated users do not understand how to use the tables and, as a result, tend not to use them. Others want only an easy procedure to determine the releasability of an estimate: releasable, qualified, or not releasable. Sophisticated users find the tables neither detailed enough nor adequate for complex analyses such as linear or logistic regression analyses. Still others find them burdensome because it is a manual procedure and they have many estimates for which they require CV's. In summary, it appears that there are two distinct groups of users:

1. Basic analysts for whom the CV look-up tables are appropriate (if and when they use them),

2. Sophisticated analysts who find the CV look-up tables burdensome and/or inadequate.

For the first group of users, it would be desirable to

find a more automated method to obtain the approximate CV's, while for the second group a method for the analyst to calculate a design valid variance estimator is desired.

3. Alternative Methods

At Statistics Canada, two alternative methods have been investigated as solutions to the PUMF variance estimation problem. The first approach utilizes the Generalized Variance Function (GVF) approach of Wolter (1985) while the second approach uses the jackknife variance estimator for a stratified multi-stage design in conjunction with collapsing of design strata and clusters.

3.1 Generalized Variance Functions

The method of generalized variance functions uses a mathematical model to describe the squared CV of a survey estimator and its expectation. Possible models include

$$CV^2 = \alpha + \beta/X,$$

$$CV^2 = \alpha + \beta/X + \gamma/X^2$$

and

$$CV^2 = \left(\alpha + \beta X + \gamma X^2\right)^{-1}.$$

Wolter (1985) notes that there is very little theoretical justification for any of the models presented above. For more on the choice of models, the reader is referred to Wolter (1985). To obtain estimates of α, β and γ, one would calculate many estimates of survey variables, $\hat{X}_i$, and their corresponding CV's and then using ordinary or weighted least squares, obtain the estimates $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\gamma}$. Assuming the first model was chosen, an estimate of the CV of a survey statistic $\hat{Y}$ is then obtained as

$$\hat{C}V^2 = \hat{\alpha} + \hat{\beta}/\hat{Y}.$$

Although the GVF method appears to be a possible replacement for the CV look-up tables, it still suffers from some of the same problems experienced by the look-up tables. For the basic analysts, the GVF method provides an easy and somewhat automated method of obtaining approximate CV's. Unfortunately, attempts to develop GVF techniques for continuous variables have been largely unsuccessful. As well, the GVF method is not valid for complex statistics such as regression parameters. Even with these short-comings, the GVF method is in use for the U.S. Bureau of the Census' Survey of Income and Program Participation (SIPP) and has been investigated for Statistics Canada's Survey of Labour and Income Dynamics (SLID).

3.2 Collapsing

For use with Statistics Canada's National Population Health Survey (NPHS), Mayda et al. (1996) proposed using the collapsing method of Rust (1986) to create "super-strata" and "super-clusters" and then applying the usual jackknife variance estimator on the super-strata and super-clusters. Following Rust (1986), design strata are collapsed to form super-strata and then the original clusters are collapsed within the super-strata. The clusters are collapsed in such a way that the super-clusters contain original clusters from the same design strata. The super-strata and super-cluster identifiers are then included on the PUMF thus allowing analysts to use the jackknife variance estimator. This method is illustrated in Mayda et al. (1996) using data from the NPHS. Although results from their empirical study are encouraging, one should take care when collapsing strata and clusters within strata as Valliant (1995) has shown that under certain conditions the balanced repeated replication (BRR) variance estimator can become inconsistent when strata are collapsed. It is unclear at this point whether the inconsistency property of the BRR extends to the jackknife variance estimator, but the asymptotic equivalent of the BRR and the jackknife variance estimators, as shown in Rao and Wu (1988), indicates that the jackknife variance estimator may also suffer under collapsing.

4. Bootstrap Variance Estimation

The bootstrap variance estimation method for the *iid* case has been extensively studied, see Efron (1982). Rao and Wu (1988) provided an extension to stratified multi-stage designs but covered only smooth statistics $\hat{\theta} = g(\hat{Y})$. The design considered by Rao and Wu (1988), and in this paper, assumes L design strata with N_h clusters in the hth stratum. Within the hth stratum, $n_h (\geq 2)$ clusters are selected and further subsampling within selected clusters is performed according to some probability sampling design. Although the subsampling is not specified, it is assumed that there is unbiased estimation of cluster totals, $Y_{hi}, h = 1, \ldots, L; i = 1, \ldots, n_h$. From the survey design, design weights, w_{hik}, associated with the $(hik)^{th}$ sampled element are obtained. Also associated with the $(hik)^{th}$ sampled element is the variable of interest, y_{hik}. An estimator of the total Y is given by

$$\hat{Y} = \sum_{(hik) \in s} w_{hik} y_{hik} \qquad (4.1)$$

where s denotes the sampled elements.

Under this design, Rao, Wu and Yue (1992) generalized the Rao-Wu bootstrap to cover non-smooth statistics as well as smooth statistics. To calculate a bootstrap variance estimator for $\hat{\theta} = g(\hat{Y})$, where $\hat{Y}$ is given by equation (4.1), the Rao-Wu-Yue method is as follows:

i) Independently, in each stratum, select a simple random sample of m_h clusters with replacement from the n_h sample clusters.

ii) Let m_{hi}^* be the number of times the $(hi)^{th}$ cluster is selected $(\sum_i m_{hi}^* = m_h)$. Define the bootstrap weights as

$$ w_{hik}^* = \left[\left\{ 1 - \left(\frac{m_h}{n_h - 1} \right)^{1/2} \right\} + \quad (4.2) \right.$$
$$ \left. \left(\frac{m_h}{n_h - 1} \right)^{1/2} \frac{n_h}{m_h} m_{hi}^* \right] w_{hik}.$$

If the size of the simple random sample, m_h, is chosen to be less than or equal to $n_h - 1$, then the bootstrap weights, w_{hik}^*, will all be positive.

iii) Calculate $\hat{\theta}^*$, the bootstrap estimator of θ, using the bootstrap weights, w_{hik}^*, in the formula for $\hat{\theta}$.

iv) Independently replicate steps (i) to (iii) a large number of times, B, and calculate the corresponding estimates, $\hat{\theta}_{(1)}^*, \ldots, \hat{\theta}_{(B)}^*$.

The bootstrap variance estimator for $\hat{\theta}$ is then given by

$$ v_B(\hat{\theta}) = \frac{1}{B} \sum_b \left(\hat{\theta}_{(b)}^* - \hat{\theta}_{(\cdot)}^* \right)^2 $$

where $\hat{\theta}_{(\cdot)}^* = (1/B) \sum_b \hat{\theta}_{(b)}^*$.

While Rao, Wu and Yue (1992) have considered only the design weights, w_{hik}, they note that the design weights are often subjected to adjustments such as poststratification or generalized regression to ensure consistency to known population totals. For example, suppose population totals for age groups and race groups are known but not the individual age-race group totals. In terms of a two-way table, the marginal totals are known but not the individual cell totals. The generalized regression (GREG) estimator can use the known population totals to increase the efficiency of estimates (see Huang and Fuller 1978; Deville and Särndal 1992).

To define the GREG estimator, let $\mathbf{x}_{hik}$ be a vector of auxiliary variables with known population totals $\mathbf{X}$. The GREG estimator of Y is given by

$$ \hat{Y}_R = \sum_{(hik) \in s} \tilde{w}_{hik} y_{hik} $$

where $\tilde{w}_{hik} = w_{hik} a_{hik}$ are the regression adjusted weights, $a_{hik} = 1 + \mathbf{x}_{hik}^T \hat{\mathbf{A}}^{-1} (\mathbf{X} - \hat{\mathbf{X}})$ is the regression adjustment, $\hat{\mathbf{X}} = \sum_{(hik) \in s} w_{hik} \mathbf{x}_{hik}$ and $\hat{\mathbf{A}} = \sum_{(hik) \in s} w_{hik} \mathbf{x}_{hik} \mathbf{x}_{hik}^T$. To calculate a bootstrap variance estimator for $\hat{Y}_R$, the following step is added to the bootstrap method described above:

iia) To obtain the bootstrap final weights, perform the same weight adjustment with the design weights, w_{hik}, replaced by the bootstrap weights, w_{hik}^*. For example, under a regression adjustment, the bootstrap final weights are given as

$$ \tilde{w}_{hik}^* = w_{hik}^* a_{hik}^* \qquad (4.3) $$

where

$$ a_{hik}^* = 1 + \mathbf{x}_{hik}^T \tilde{\mathbf{A}}^{-1} (\mathbf{X} - \hat{\mathbf{X}}^*), \quad (4.4) $$
$$ \hat{\mathbf{X}}^* = \sum_{(hik) \in s} w_{hik}^* \mathbf{x}_{hik} \qquad (4.5) $$

and

$$ \tilde{\mathbf{A}} = \sum_{(hik) \in s} w_{hik}^* \mathbf{x}_{hik} \mathbf{x}_{hik}^T. \qquad (4.6) $$

These bootstrap final weights are then used in place of the bootstrap weights in step (iii).

A commonly used value for m_h is $n_h - 1$, in which case equation (4.2) reduces to

$$ w_{hik}^* = \frac{n_h}{n_h - 1} m_{hi}^* w_{hik}. \qquad (4.7) $$

If a sampled element is in a cluster that has not been selected in a particular bootstrap sample, then $m_{hi}^* = 0$ and the bootstrap final weight is equal to zero. That is, all sampled elements in the cluster have bootstrap weights equal to zero and in the case of multiplicative weight adjustments (ex. poststratification or regression), will also have bootstrap final weights equal to zero. Now, within each bootstrap sample at least one cluster per stratum will have bootstrap final weights equal to zero, so that members of the same cluster can not be identified by their zero weights. Unfortunately, when the bootstrap final weights are combined over all B bootstrap samples, cluster membership can be identified. By grouping individuals based on zero and non-zero

bootstrap final weights, the members of each cluster can be identified. Although location of the clusters is not given on the PUMF, use of other variables on the PUMF may allow users to deduce the location of a cluster and thus breaching confidentiality.

This problem occurs because under a stratified multi-stage design, the bootstrap resamples entire clusters. In the case of stratified simple random sampling, confidentiality is conserved since the cluster consists of a single element. Unfortunately, for stratified multi-stage samples (commonly used in social surveys), the bootstrap method does not meet Statistics Canada's confidentiality guidelines.

As a possible solution to this problem, it was suggested to change the size of the simple random sample, m_h, so that equation (4.2) does not reduce to (4.7). Reducing m_h to be less than $n_h - 1$ may cause problems as it is common to select only 2 clusters per stratum. In this case, some sort of collapsing would be necessary to increase the number of clusters per stratum. Increasing m_h to be greater than $n_h - 1$ would lead to negative bootstrap weights, which does not pose any problems as long as the analysts accept the negative weights and use them only for variance estimation and not for point estimation. Upon closer examination though, it was noted that the zero weights obtained by using $m_h = n_h - 1$ were simply replaced by negative weights and the problem with confidentiality still exists.

Two solutions have been investigated to solve the confidentiality problem: modifying the regression adjustment and replacing the bootstrap weight with an average bootstrap weight. The two solutions will be presented separately.

4.1 Modified Regression Adjustment

In the context of jackknife variance estimation for the GREG estimator, Yung and Rao (1996) state that the jackknife variance estimator can be approximated by retaining the inverse for the full sample, $\hat{\mathbf{A}}^{-1}$, and modifying the regression adjustment a_{hik}. They showed that the resulting jackknife variance estimator is equivalent to the standard linearization variance estimator and is therefore a consistent estimator of the variance of the GREG estimator. In the context of the bootstrap, the modified regression adjustment for $w_{hik}^* \neq 0$ is

$$\tilde{a}_{hik}^* = 1 + (w_{hik}/w_{hik}^*)\mathbf{x}_{hik}^T\hat{\mathbf{A}}^{-1}(\mathbf{X} - \hat{\mathbf{X}}^*)$$

and the modified bootstrap final weight is

$$\tilde{w}_{hik}^* = w_{hik}^*\tilde{a}_{hik}^* \text{ if } w_{hik}^* \neq 0$$

and

$$\tilde{w}_{hik}^* = w_{hik}\mathbf{x}_{hik}^T\hat{\mathbf{A}}^{-1}(\mathbf{X} - \hat{\mathbf{X}}^*) \text{ if } w_{hik}^* = 0.$$

That is,

$$\tilde{w}_{hik}^* = w_{hik}^* + w_{hik}\mathbf{x}_{hik}^T\hat{\mathbf{A}}^{-1}(\mathbf{X} - \hat{\mathbf{X}}^*).$$

Thus, even if $w_{hik}^* = 0$, the final bootstrap weight is equal to

$$\tilde{w}_{hik}^* = w_{hik}\mathbf{x}_{hik}^T\hat{\mathbf{A}}^{-1}(\mathbf{X} - \hat{\mathbf{X}}^*)$$

which is zero only if $w_{hik} = 0$ or $\mathbf{X} = \hat{\mathbf{X}}^*$. The use of this modified regression adjustment will not affect the accuracy of the bootstrap variance estimator since the term $\mathbf{X} - \hat{\mathbf{X}}^*$ converges to zero asymptotically, but the resulting bootstrap variance estimator maybe slightly less efficient.

4.2 Mean Bootstrap Weights

The confidentiality problem occurs when $m_{hi}^* = 0$ for one or more clusters. To avoid this problem, produce R bootstrap samples and average the m_{hi}^*'s over the R samples. As long as each cluster appears in at least one of the R bootstrap samples, the averages will all be non-zero. The steps to perform the mean bootstrap are as follows:

i) Independently, in each stratum, select a simple random sample of $(n_h - 1)$ clusters with replacement from the n_h sample clusters.

ii) Repeat step (i) R times.

iii) Let $m_{hi(r)}^*$ be the number of times the $(hi)^{th}$ cluster is selected in the rth bootstrap sample. Let $m_{hi(\cdot)}^* = (1/R)\sum_r m_{hi(r)}^*$ be the average number of times the $(hi)^{th}$ cluster is selected over the R bootstrap samples.

iv) Define the mean bootstrap weight as

$$w_{hik(\cdot)}^* = \frac{n_h}{n_h - 1}m_{hi(\cdot)}^*w_{hik}.$$

v) Obtain the bootstrap final weights, $\tilde{w}_{hik(\cdot)}^*$, by performing the same weight adjustment except replace the design weight, w_{hik}, with the mean bootstrap weight, $w_{hik(\cdot)}^*$. For example, in the context of a regression adjustment, we simply replace w_{hik}^* in equations (4.3), (4.4), (4.5) and (4.6) with $w_{hik(\cdot)}^*$.

vi) Calculate $\tilde{\theta}^*$ using the bootstrap final weights in the formula for $\hat{\theta}$.

vii) Independently replicate steps (i) to (vi) a large number of times, B, and calculate the corresponding estimates, $\tilde{\theta}^*_{(1)}, \ldots, \tilde{\theta}^*_{(B)}$.

The bootstrap variance estimator is then given as

$$v_{MB}(\hat{\theta}) = \frac{R}{B} \sum_b (\tilde{\theta}^*_{(b)} - \tilde{\theta}^*_{(\cdot)})^2$$

where $\tilde{\theta}^*_{(\cdot)} = (1/B) \sum_b \tilde{\theta}^*_{(b)}$. Heuristically, one can view $\tilde{\theta}^*_{(b)}$ as an estimate of a mean so that $(1/B) \sum_b (\tilde{\theta}^*_{(b)} - \tilde{\theta}^*_{(\cdot)})^2$ is an estimator for $(1/R)Var(\hat{\theta})$, thus justifying the factor of R in the formula for $v_{MB}(\hat{\theta})$. The size of R should be large enough so that the chance of $m_{hi(m)} = 0$ for all $r = 1, \ldots, R$ is very small, but it should not be so large that drawing $R \times B$ bootstrap samples becomes unrealistic.

5. Empirical Comparisons

Although consistency of the bootstrap variance estimator has been previously established, the performance of the bootstrap variance estimator was compared to the CV look-up tables empirically. For this comparison, the PUMF of Statistics Canada's 1994 National Population Health Survey (NPHS) was used.

5.1 National Population Health Survey

The NPHS was designed to collect information related to the health of the Canadian population. The objectives of the NPHS included:

- to aid in the development of public policy by providing measures of the level, trend and distribution of the health status of the population;

- to provide data for analytic studies that will assist in understanding the determinants of health;

- to increase the understanding of the relationship between health status and health care utilization, including alternative as well as traditional services.

The design of the NPHS consisted of a stratified two-stage design. In the first stage, homogeneous strata were formed and independent samples of clusters were drawn from each stratum. In the second stage, dwellings were selected within each sampled cluster. The design weights were obtained based on both stages of sampling. To obtain the final weight, a series of twelve weighting adjustments were performed with the last adjustment being a poststratification adjustment. More information on the NPHS design and weighting is available in the NPHS PUMF documentation.

To implement the bootstrap, $n_h - 1$ clusters were sampled with replacement within each stratum. Mean bootstrap weights were calculated using twenty bootstrap samples and 100 sets of mean bootstrap weights were generated (i.e. 2000 bootstrap samples in total). For each set of mean bootstrap weights, only the poststratification adjustment was performed. For comparison purposes, the full jackknife variance estimator, without any collapsing, was used to obtain a 'true CV'. The choice of the jackknife CV as the true CV is justified since if an analyst wishes to obtain a true CV, the jackknife variance estimator would be used at Statistics Canada to calculate the CV. Table 5.1 gives some results of the comparison between the CV's obtained from the bootstrap, collapsed jackknife and CV look-up tables for means, totals and ratios of categorical and continuous variables.

Table 5.1 - National Population Health Survey

$CV_i - CV_J$	Bootstrap	Collapsed Jackknife	CV Table
± 1 %	64	41	45
	(86.8 %)	(45.6 %)	(50.0 %)
± 2 %	82	62	58
	(91.1 %)	(68.9 %)	(64.4 %)
± 3 %	88	72	64
	(97.8 %)	(80.0 %)	(71.1 %)
± 4 %	90	77	71
	(100.0 %)	(85.6 %)	(78.9 %)
> 4 %		90	75
		(100.0 %)	(83.3 %)

In Table 5.1, CV_i is the CV based on the bootstrap, collapsed jackknife or the look-up tables and CV_J is the CV obtained from full jackknife variance estimator. Table 5.1 shows one of the drawbacks of the CV look-up table method. Of the 90 estimates, 15 involved a continuous variable which could not be handled by the CV look-up tables. Of the remaining 75 estimates, 71 were within $\pm 4\%$ of the true CV. The bootstrap and the collapsed jackknife both performed well for all estimates with the bootstrap performing better than the collapsed jackknife (87% versus 45% of the estimates were within $\pm 1\%$ and 100% versus 85% of the estimates were within $\pm 4\%$).

To illustrate the versatility of the bootstrap method, CV's were calculated for the regression coefficients for the model relating a person's health status to a measure of the restriction of activities, age, type of drinker and household income. The model was fit separately within five provinces giving a total of 25 parameter estimates. Again, the full jackknife CV was treated as the true CV. The results of this comparison are given in Table 5.2

Table 5.2 - Regression Coefficients for NPHS

$CV_i - CV$	Bootstrap	Collapsed Jackknife
±1 %	13 (52.0 %)	6 (24.0 %)
±2 %	19 (76.0 %)	11 (44.0 %)
±3 %	20 (80.0 %)	15 (60.0 %)
±4 %	21 (84.0 %)	18 (72.0 %)
> 4 %	25 (100.0 %)	25 (100.0 %)

From Table 5.2, we can see that the bootstrap performs better than the collapsed jackknife with 52% of the estimated CV's within ±1% for the bootstrap compared with only 24 % for the collapsed jackknife. Also, the bootstrap has slightly more estimates within ±4% (84 % versus 72 %).

6. Conclusions

The addition of bootstrap final weights to PUMF's will allow users to calculate correct design-based variance estimators (and hence CV's) for categorical and continuous variables as well as complex statistics such as regression coefficients. Although some technical knowledge is required, it is felt that the calculation of the bootstrap variance estimators is straight forward. Empirical comparisons with the currently used CV look-up method shows the superiority of the bootstrap method both in terms of accuracy and the types of estimators which can be used. While comparisons with the collapsed jackknife indicate only a slightly better performance for the bootstrap CV's, it is felt that the bootstrap methodology has better theoretical justifications.

7. References

Deville, J.C., and Särndal, C.E. (1992). Calibration Estimators in Survey Sampling, *Journal of the American Statistical Association*, **87**, 376-383.

Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans*. Philadelphia: Society for Industrial and Applied Mathematics.

Huang, E.T., and Fuller, W.A. (1978). Nonnegative Regression Estimation for Sample Survey Data, in *Proceedings of the Social Statistics Section*, Washington, D.C.: American Statistical Association, pp. 300-305.

Mayda, J.E., Mohl, C., and Tambay, J.-L. (1996). Variance Estimation and Confidentiality: They are Related!, in *Proceedings of the Survey Methods Section*, Statistical Society of Canada, pp. 135-141.

National Population Health Survey Public Use Microdata File Documentation, Statistics Canada Publication Number 82F0001XCB

Rao, J.N.K., and Wu, C.F.J. (1988). Resampling Inference with Complex Survey Data. *Journal of the American Statistical Association*, **83**, 231-241.

Rao, J.N.K., Wu, C.F.J., and Yue, K. (1992). Some Recent Work on Resampling Methods for Complex Surveys. *Survey Methodology*, **18**, 209-217.

Rust, K. (1986). Efficient Replicated Variance Estimation, in *Proceedings of the Survey Research Methods Section*, Washington, D.C.: American Statistical Association, pp. 81-87.

Valliant, R. (1995). Limitations of Balanced Half Sampling when Strata are Grouped, in *Proceedings of the Survey Research Methods Section*, Washington, D.C.: American Statistical Association, pp. 120-125.

Wolter, K.M. (1985). Introduction to Variance Estimation. Springer-Verlag New York Inc.

Yung, W., and Rao, J.N.K. (1996). Jackknife Linearization Variance Estimators Under Stratified Multi-Stage Sampling. *Survey Methodology*, **22**, 23-31.