

The Replication Method for Estimating Sampling Errors, Principles, and Software

David Morganstein and J. Michael Brick, Westat, Inc.

Abstract

Replication is a very useful method of approximating the variance of nonlinear statistics in complex sample surveys. It provides asymptotically equivalent results to the alternative linearization method and it has distinct strengths under some very practical situations that commonly arise in sample surveys. In this paper, we describe how replication theory can be implemented in software such as WesVarPC to achieve some important objectives that are inherently difficult with linearization methods. The strengths of replication and WesVarPC include: it is easy to explain replication to data users without special training in variance estimation for sample surveys; it can be applied to nearly all statistics using the same procedure without modifying the software; it can handle a variety of estimation strategies; it is a consistent procedure for handling subsets of the data and missing data items, and it is a simple and elegant method for handling longitudinal sample designs

1. Introduction

In a sample survey, a common goal is to estimate a population quantity, $\theta = g(y_1, \dots, y_N)$, using a statistic, $\hat{\theta} = g(y_1, \dots, y_n)$, where the y_i may be vectors. The design that gives rise to the sample might be complex, involving unequal probabilities of selection, clustering, stratification, and several stages of sampling. In such circumstances, the design must be taken into account to produce unbiased estimates and standard errors of the estimates. Most standard statistical software such as SPSS and the SAS® System ignore the sample design and assume the observations are independent and identically distributed. Using weighting features of the standard software can produce unbiased point estimates, but it is not possible to correctly estimate standard errors. In fact, if either the sample design or the estimator is complex, the exact expression for the variance of the estimator is often intractable and approximate methods are needed.

One approach to approximating the variance of an estimate is to use replication methods which include the jackknife, balance repeated replication (BRR), and bootstrap methods. For the most part, in this article we only discuss jackknife and BRR options to illustrate the main points. An alternative approach is to linearize the estimator using a Taylor series expansion and then use

standard sample survey variance estimation methods to estimate the precision of the linearized statistic. Both the linearization and replication methods are described in more detail in Wolter (1985).

Many theoretical and empirical studies compare replication and linearization methods of variance estimation. On the theoretical side, early work by Krewski and Rao (1981) show the asymptotic equivalence of these methods. Subsequent theoretical work, including that of Rao, Wu and Yue (1992), show that the linearization and replication methods are asymptotically equivalent to a reasonably high order. On the empirical side, simulation studies have been conducted to compare the precision of the estimates using both methods. One of the earliest of these studies is Kish and Frankel (1974). More recent work by Kovar, Rao and Wu (1988), Rao, Wu, and Yue (1992), and Valliant (1990) show that the methods generally result in consistent estimates. The jackknife's properties are very similar to those of linearization, while the properties of the BRR and bootstrap methods are very similar to each other. Most of the comparisons consider both the accuracy of the variance estimates and the derived confidence intervals. Rust and Rao (1996) review some of this literature and examine other aspects of the two approaches to approximating variances of estimates from complex sample surveys.

In this paper, we examine circumstances in which the replication method has some distinct advantages over the linearization method for estimating variances in sample surveys. Since both methods result in estimates that are asymptotically equivalent, the criteria for assessing the strengths of the replication method are factors other than the precision of the estimates. Some of the features examined are the survey design and estimation strategies that can be more easily reflected appropriately with replication than linearization. Of course, in some situations linearization is simpler or easier to implement but most of these situations are covered in other literature (Shah et al., 1992).

The strengths of a method depend heavily on the software that is available for use in computing. WesVarPC, a Windows-based software package for computing variances using replication methods, is used to illustrate the strengths of the replication method in practical applications. WesVarPC is designed to compute functions of weighted totals and fit models to data from complex samples using a variety of replication techniques

(Westat, 1997). WesVarPC is available free-of-charge from the World Wide Web URL of Westat (<http://www.westat.com>). The software and its documentation can be downloaded from this site.

In the next section, we provide a brief introduction to replication theory and methods. Particular features of sample surveys that may be more easily and accurately addressed using replication methods are discussed subsequently, with examples of how this can be accomplished illustrated using WesVarPC.

2. Review of Replication Theory

The basic idea underlying replication is to select subsamples repeatedly from the whole sample, to calculate the statistic of interest for each of these subsamples, and then to use the variability among these subsample or replicate statistics to estimate the variance of the full sample statistics. There are different ways of creating subsamples from the full sample. The subsamples are called replicates and the statistics calculated from these replicates are called replicate estimates.

Replicate variance estimators, $v(\hat{\theta})$, take the form

$$v(\hat{\theta}) = c \sum_{k=1}^G h_k (\hat{\theta}_{(k)} - \hat{\theta})^2 \quad (1)$$

where

- θ is an arbitrary parameter of interest,
- $\hat{\theta}$ is the estimate of θ based on the full sample,
- $\hat{\theta}_{(k)}$ is the k -th estimate of θ based on the observations included in the k -th replicate,
- G is the number of replicates,
- c is a constant that depends on the replication method, and
- h_k is a stratum specific constant that is only required for certain designs.

The value of c depends on the replication method; for BRR and bootstrap $c=1/G$; for the unstratified jackknife, $c=(G-1)/G$, for the paired jackknife $c=1$. The factor h_k is needed for jackknife designs with an unequal number of PSUs per stratum and can also be used to adjust for finite population corrections in some designs. This issue is not discussed further in this paper. Rust and Rao (1996) describe replication methods more completely and suggest how these methods might be appropriate for specific sample designs.

WesVarPC is designed to accommodate these and other replicate methods. An appendix to the documentation describes the specific replication methods that can be developed within WesVarPC and also describes how users can use replicate weights they have already created. Designing the replicates is a critical feature that has important implications for the validity of the estimates of the variances. For some designs, replicate preparation requires considerable ingenuity so that precise variance estimates can be produced while not requiring too much computation. For other designs, it is very simple.

Before we turn our attention to circumstances in which replication has advantages over linearization methods, we examine some of the key features of WesVarPC. This is necessary because the software is the mechanism for taking advantage of the method and no benefit is gained unless the software makes it computationally feasible.

3. Implementing Replication Methods in WesVarPC

The data handling (in particular the input and output of files) and analysis features of WesVarPC are reviewed here to give some context to the use of the software with replication designs. Data collected from complex surveys are often stored in SPSS, SAS or database files to facilitate editing and other data preparation activities. WesVarPC can read all of these types of files (only older version of SAS and SAS transport files can be read). If replicate weights are already on these files, they can be read directly into WesVarPC. Otherwise, WesVarPC has procedures for creating jackknife and BRR replicate weights. Furthermore, the replicate weights can be adjusted to known population totals using poststratification or raking procedures within the program. After these replicate weights are created once, they are retained for all future analysis. The output of the software can be written to a flat file, but newer versions of the software are planned that will provide much greater control over the output using standard Windows 95 structures.

There are two analytic functions that produce estimates and their standard errors. One is for tables and the other for regression models. These features are much more extensive than this brief statement suggests. Users can specify multi-dimensional tables and automatically obtain estimates of totals (sum of the weights), percentages, and weighted totals (for continuous variables, such as income) for each cell in the cross-classification. It is also simple to define statistics that are ratios, differences, or more complex functions of statistics using the "Compute" statistics program interface (a simple screen is provided for this). In addition, complex functions of the estimates in cells of the table can be defined as "Function" statistics using an interface similar to that for Compute statistics. Examples of Function

statistics that users can define include linear contrasts and log-odds ratios. Other table features, such as adjusted chi-square test statistics are too numerous to mention here.

Linear and logistic regression analysis are also supported in WesVarPC. These two methods encompass many other statistical procedures such as analysis of variance and covariance as special cases. A simple interface is used to specify the dependent variable and independent variables, including interaction terms. Estimates of the model parameters, their standard errors, and statistical tests of hypotheses about these parameters are produced by WesVarPC.

This discussion provides only the briefest background on the software. Some of these issues are discussed further below. The WesVarPC User Manual describes all the features of the software.

4. Some Strengths of Replication

4.1 Ease of Use

The choice of which method of variance estimation to use can be examined in a number of ways. One important dimension is the ability to describe how the variance estimates are produced to users without advanced training in variance estimation for sample surveys. Replication is a very well-known method in statistics and many users readily accept the basic premise of replication. Dividing the full sample into subsamples and estimating the variability of replicate subsamples around the full sample estimate is often the only explanation required.

This aspect of ease of use is re-enforced by the fact that the same procedure is followed for nearly all statistics: means, totals, ratios, and more complex functions of the sample data. For example, the log-odds ratio for a two-way table from a sample can be written as:

$$\hat{\theta} = \log\left(\frac{\hat{N}_{11}\hat{N}_{22}}{\hat{N}_{12}\hat{N}_{21}}\right) \quad (2)$$

where the '^' above the quantities indicate these are consistent, approximately unbiased sample estimates of the population characteristics. The variance of the sample log-odds ratio can be estimated exactly the same as the sample mean or total using (1).

This simplicity contrasts with the linearization method that requires each nonlinear statistic be linearized before the variance can be estimated. The linearized form of the statistic must be developed and incorporated into the software. While the derivation of the linearized form is often not difficult (Binder, 1996), the linearization

software must either have the specific statistic already explicitly included in its list of statistics or it must numerically approximate the statistic. Either of these options places some limits on the data user.

A very important strength of replication is that it allows the user to define nearly any statistic and compute the variance of it using (1) without having to modify the software. WesVarPC does this in two ways. First, the user can specify a "Compute" statistic as a function of variables on the data set, as described above. For example, some "Compute" statistics that are easily defined in WesVarPC are ratios of ratios, the log of ratios, the square root of a difference of ratios. An easy to use graphical interface is used to enter the statistic and the program computes (1) for that statistic. Second, the user can specify a "Function" statistic as a function of estimated quantities in the cells of a table. For example, the log-odds ratio example mentioned above can be estimated in WesVarPC by naming the four cells of the two-by-two table and specifying the log of the ratio of the estimated totals in the cells. User specified statistics, such as differences, contrasts, and logs of differences or ratios can be calculated this way. Once again, the simplicity is possible because the variance is estimated using (1) for almost all types of statistics.

4.2 Handling Different Estimation Methods

In most sample surveys, it is not sufficient to simply weight the observed data by the reciprocal of the probability of selection. Different estimation strategies are used to reduce the bias and/or the variance of the estimates. For example, almost all surveys are subject to nonresponse and the weights are adjusted to compensate for this nonresponse (Brick and Kalton, 1996). Similarly, estimation strategies such as poststratification, raking, or generalized regression estimation are often used to take advantage of auxiliary data and reduce both the bias and the variances of the estimates. Replication offers a simple and elegant method of incorporating these adjustments in the estimates of variance (Yung, 1996).

Rust and Rao (1996) show that a simple way of creating the replicate estimates required in (1) is to create replicate weights and attach these to each observation just as the full sample weight is normally attached to an observation. The replicate estimates can then be computed by using the replicate weights the same way the full sample weights are used to create the full sample estimate. They also point out that, if the agency responsible for data collection includes the replicate weights with the data, this results in three important advantages: first, the user does not have to know any of the features of the sample design to properly estimate sample variances (the information is all contained in the replicate weights); second, the agency can reflect all stages of estimation

appropriately in the replicate weights so as to reflect the full estimation scheme used in the full sample weights; and, third, the software for replicate variance estimation does not have to be modified to compute the estimates properly (equation (1) is still used).

WesVarPC allows users to input replicate weights that have been adjusted for nonresponse, ratio estimation, poststratification, raking, calibration estimation, or any other commonly used method of estimation. Thus, the ordinary user can take advantage of the advanced estimation procedures used by the agency collecting the data to reduce the bias and the variance of the estimates, provided the agency places these replicate weights on the data file. For example, if nonresponse adjustments are used in creating the final weights, these same nonresponse adjustments can be applied separately to each replicate (each replicate thus has a different nonresponse adjustment) to reflect this procedure in the computed estimates. In addition, WesVarPC provides an option to allow the user to create replicate weights with poststratification or raking adjustments if the control totals required for these procedures are available and the replicate weights do not already include these adjustments.

This simplicity contrasts with the linearization approach, where each different estimation strategy requires a different development. For example, poststratification can be handled in linearization theory as a type of ratio estimation, but this implies that the estimator is different for the poststratified and the simple Horvitz-Thompson type of estimator. Thus, linearization software requires that the specific linearized poststratified estimator be included in the software to compute the variance appropriately. Nonresponse adjustments could also be handled in the linearization approach, but again it places a severe burden on the software to accommodate each type of estimation scheme. We are not aware of any linearization software that handles nonresponse adjusted weights in a simple manner. In replication, this is very simple if the agency providing the data includes this component in the replicate weights. The process to do this is straight-forward and needs to be done only once when the weights are created.

4.3 Handling Subsets and Missing Data

Since the replicate weights carry all the information needed for estimating variances, replication is also extremely well-suited for handling data that are subset or missing. For example, a sample of adults might have been selected, but some analysts may wish to only estimate characteristics for a subset of the adults, say females between the ages of 30 and 44. This type of analysis is simple to do with replication methods without concerns about the implications for variance estimation.

For example, in WesVarPC the subset of the data can be identified using a simple subsetting statement and a subset file is created that contains only the records for the included adults, along with their full sample weight and replicate weights. The detailed analysis of just these records can then be conducted with the subset file since the replicate weights contain all the information needed to properly estimate the precision of the estimates.

This process is not as simple in most linearization software because the design information may no longer be accurate if the subset file is used alone. The basic problem is that the subset may not include observations from all of the sampled primary sampling units and most of the linearization software requires that all these units be identified in the data file used for analysis. Graubard and Korn (1996) show what might happen in this situation and illustrate the consequences with two examples from actual surveys. Here again, it is feasible to handle this situation under the linearization approach, but the steps required are more complex and more susceptible to error than the simple replication approach.

Another circumstance that is common in sample surveys but difficult to handle is missing data. Some of the issues associated with handling nonresponse adjustments of the weights that are normally used to adjust for unit nonresponse were mentioned earlier. Now, we turn our attention to item nonresponse and the missing data from this source.

Handling missing items in analysis is a long-standing problem without a fully accepted, general purpose, solution. It is too ambitious to expect to handle missing data in variance estimation to everyone's satisfaction when a method has not been accepted for handling the estimates themselves. However, a more modest goal is to handle the missing items in a consistent manner for creating the estimate and for estimating the variance of the estimates. In this regard, replication has a decided advantage over linearization methods.

Once a procedure has been determined for handling missing data for creating an estimate, that same procedure can be applied to create both the full sample and the replicate estimates. WesVarPC does this; the full sample and replicate weights are used in exactly the same way with respect to missing data for creating the components needed for computing (1). For example, in estimating a total with missing item responses, the full sample estimate and the replicate estimates may both be biased downward because of item nonresponse (if no imputation is done for the item), but the variance of the estimate is computed appropriately for the procedure.

In the linearization approach, this type of consistency is more difficult to achieve. In fact, many of the

linearization software packages require all the items have responses in order to be able to compute variances. Other linearization software allow missing data, but do it in ways that are not consistent because the variance depends on the sample design and the specific form of the estimator. For example, the problem mentioned above with subsetting may also occur due to missing item responses but it may be very hard to detect that this has occurred. The methods of handling missing data also may result in estimates for strata being computed differently than estimates from the full sample.

The implications for handling missing data can have important consequences. For example, the software may compute estimates of totals by the equivalent of imputing a value of zero for the missing item, while means may be computed by ignoring the observations with missing data. The result is that different statistics, means and totals, are not treated the same. An example shows how important this can be. Suppose there are 4 respondents, 3 reporting values of 99, 100, and 101, while 1 has a missing value. In WesVarPC and replication methods, the estimated total for the full sample would be 300 and each replicate estimate would be nearly equal to 300 so that the estimated variance would be very small. On the other hand, if the linearization program imputed a value of zero for the missing value, then the estimated variance would be very large (the 4 observations for computing the variance would be 99, 100, 101, and 0).

4.4 Handling Longitudinal or Panel Survey Designs

Panel or longitudinal sample surveys have become very popular in the last several decades, especially because they offer the possibility of measuring change over time more precisely than independent cross-sectional surveys. The increased precision is the result of the correlation between the units sampled over time. However, some of the benefits of having a more precise estimate cannot be realized unless the estimates of precision reflect the correlated responses.

Replication methods provide a rather simple and elegant method for incorporating the correlation. The replicates defined for units sampled for the baseline sample can be used also for the follow-up data collections so that the replicate estimates defined in (1) naturally contain the appropriate overlap and thus the estimates of precision reflect the design. Hinkins et al. (1996) discuss this issue in more detail. Once again, WesVarPC can be used without modification to produce appropriate estimates of precision for estimates, provided the replicate weights for the follow-up design are produced correctly.

For linearization software, the problem is more complex. For some estimates it is possible to construct new variables that are the difference between the data

collected in the different waves of the survey, but this may involve additional burdens on the data users. In some cases, elaborate procedures are required to address this issue.

References

- Binder, D.A. (1996) Linearization methods for single phase and two-phase samples: a cookbook approach. *Survey Methodology*, 22: 17-22.
- Brick, J.M., Kalton, G. (1996) Handling missing data in survey research. *Statistical Methods in Medical Research*; 5: 215-38.
- Graubard, B.I., Korn, E.L. (1996) Survey Inference for Subpopulations, *American Journal of Epidemiology*, 144: 102-106.
- Hinkins, S., Moriarity, C., Scheuren, F. (1996) Replicate variance estimation in stratified sampling with permanent random numbers. Proceedings of the Section of Survey Research Methods of the American Statistical Association.
- Kish, L., Frankel, M. (1974) Inference from complex samples. *Journal of the Royal Statistical Society*; B36, 1-22.
- Kovar, J.G., Rao, J.N.K., Wu, C.F.J. (1988) Bootstrap and other methods to measure errors in survey estimates. *Canadian Journal of Statistics*; 16: 25-46.
- Krewski, D., Rao, J.N.K. (1981) Inference from stratified samples: properties of linearization, jackknife and balanced repeated replication methods. *Annals of Statistics*; 9: 1010-19.
- Rao J.N.K., Wu, C.F.J., Yue, K. (1992) Some recent work in resampling methods. *Survey Methodology*; 18: 209-17
- Rust, K.F., Rao, J.N.K. (1996) Variance estimation for complex surveys using replication techniques. *Survey Methods in Medical Research*; 5: 283-310.
- Shah, B.V., Barnwell, B.G., Hunt, P.N., LaVange, L.M. (1992) *SUDAAN user's manual*. Release 6.0. Research Triangle Park, NC: Research Triangle Institute.
- Valliant, R. (1990) Comparisons of variance estimators in stratified random and systematic sampling. *Journal of Official Statistics*; 6: 115-31.
- Westat, Inc. (1997) *A user's guide to WesVarPC*, Version 2.1, Rockville, MD: Westat, Inc.
- Wolter, K. (1985) *Introduction to variance estimation*. New York: Springer.
- Yung, W. (1996) Contributions to poststratification in stratified multistage samples. PhD thesis, Carleton University.

David Morganstein, Westat, Inc., 1650 Research Blvd.
Rockville, MD 20850 - 301-251-8215
morgand1@westat.com