

A Comparison of Taylor Series and JK1 Resampling Methods for Variance Estimation

Heather Hammer, Ph.D. (lead author)
Senior Study Director
Temple University Institute for Survey Research
1601 North Broad Street, Room 502
Philadelphia, PA 19122
heather@temss2.isr.temple.edu

Hee-Choon Shin, Ph.D.
Chief Statistician
Temple University Institute for Survey Research
4646 40th Street, NW – Suite 320
Washington, DC 20016
hshin@temple.edu

Lorraine E. Porcellini, M.A.
Sampling Department Manager
Temple University Institute for Survey Research
1601 North Broad Street, Room 502
Philadelphia, PA 19122
lorraine@temss2.isr.temple.edu

A Comparison of Taylor Series and JK1 Resampling Methods for Variance Estimation

Heather Hammer, Ph.D, Hee-Choon Shin, Ph.D, and Lorraine E. Porcellini, M.A.

Temple University Institute for Survey Research

ABSTRACT

It is a well-known fact in survey research that conventional software packages such as SAS and SPSS can only provide variance estimates for simple random samples, and that the naïve use of such software for variance estimation of complex survey data may lead to underestimating the variances (Weng et al. 1995). Two of the most widely used variance estimation methods for complex surveys are the jackknife (JK) method and the Taylor series method (Wolter, 1985). The purpose of this paper is to compare the variances computed by jackknife 1 (JK1) and Taylor series methods using recently released data from a large-scale survey based on a national probability sample.

KEY WORDS: Jackknife, Taylor series, variance estimation.

1. INTRODUCTION

The JK1 and Taylor series methods represent two different approaches to variance estimation. The JK1 is a resampling method, and the Taylor series is a linearization method. The advantage of the linearization method is that it is applicable to general sampling designs. However, it requires the derivation of a separate standard error formula, $s(\hat{q})$, for each nonlinear statistic, \hat{q} . In contrast, the JK1 method employs a single standard error formula for all statistics \hat{q} , and is the method of choice when explicit stratification has not been used.¹ The question addressed in this paper asks about the choice between JK1 and Taylor series methods for variance estimation in an EPSEM sample of households with clustering at the household level.

2. SAMPLE

One of the goals of the NISMART-2 Household Survey was to test for change in the incidence of missing children over time, and this was done by comparing the estimated rates of the different types of missing children based on the original NISMART-1 definitions and data collected in 1988 with those based on the original NISMART-1 definitions and the new NISMART-2 data collected in 1999 (see Hammer et al., *forthcoming*). The data used to compute the variances reported in this paper were produced by the *Adult Caretaker Household Survey* component of the Second National Incidence

¹ Note that the NISMART-2 Household Survey used systematic selection and the data have been implicitly stratified because they were sorted by state.

Studies of Missing, Abducted, Runaway, and Thrownaway Children (NISMART-2), funded by the U.S. Department of Justice, Office of Juvenile Justice and Delinquency Prevention grant number 95-MC-CX-K004. The NISMART-2 Household Survey was conducted during 1999, using a random digit dial (RDD) computer-assisted telephone interviewing methodology to collect information on missing child episodes from a national probability sample of households.

A total of 16,111 household interviews were completed with an adult caretaker, resulting in an 80% cooperation rate among eligible households with children,² and a 61% response rate.³ The total number of children identified by the survey was 31,787, and these data were weighted to reflect the census-based U.S. population of children age 18 years and younger.

The methodology used to select the NISMART-2 RDD sample for the Household Survey was a list-assisted procedure. Typically, a list-assisted RDD design is a one-stage random selection process resulting in equal probabilities of selection (EPSEM) and no clustering, with each household treated as a Primary Sampling Unit (PSU). The EPSEM design of NISMART-2 was maintained at the household level except for households with more than one telephone line, and the clustering of children within households. The effect of this clustering is illustrated by the size of the design effects⁴ reported in Table 1, all but one of which indicate a loss of precision compared to simple random sampling. Nevertheless, as indicated in Table 1, one of the design effects of interest is less than 1.0, implying that the complex design was slightly more efficient than a simple random sample with respect to this particular estimate. This may be due to the implicit stratification and systematic selection.

Although the design effects justify the use of complex variance estimation methods, this is a relatively “simple” complex design compared to complex multistage stratified and clustered designs. Therefore, one would not expect to find significant differences between the variances estimated by JK1 compared to those estimated by Taylor series methods, based on the NISMART-2 Household Survey data.

The validity of this assumption was particularly important for the Household Survey because one of the goals of NISMART-2 was to test for change in the incidence of missing children over time, and this was going to be done by comparing the estimated rates of the different types of missing children based on the NISMART-1 definitions and data collected in 1988 with those based on the original NISMART-1 definitions and the new NISMART-2 data collected in 1999 (see Hammer et al., *forthcoming*). Here, differences in the variance estimates produced by different methods could affect the results, particularly in tests where the achieved probability level was very close to the conventional level of significance selected for the test.

² American Association for Public Opinion Research (AAPOR 2000), standard definition COOP4.

³ American Association for Public Opinion Research (AAPOR 2000), standard definition RR3.

⁴ The design effect is the ratio of the variance under the actual survey design to the variance under simple random sampling with replacement.

Table 1. Design Effects for Estimated Rates of NISMART-1 Types of Missing Children Based on the 1999 NISMART-2 Data

Type of NISMART-1 Episode	Rate per 1,000 Children Age 0-17	Unweighted Number of Children	Weighted Number of Children*	JK1 Design Effect	Taylor Series Design Effect
Broad Scope Family Abduction	4.18	146	300,800	2.23	2.08
Policy Focal Family Abduction	3.15	113	227,000	1.89	1.96
Nonfamily Abduction	0.62	17	44,600	2.46	2.74
Broad Scope Runaway	5.28	154	380,400	1.08	1.51
Policy Focal Runaway	1.26	36	90,800	1.76	1.63
Broad Scope Lost, Injured, or Otherwise Missing	3.40	101	244,600	1.14	1.42
Policy Focal Lost, Injured, or Otherwise Missing	0.51	19	36,400	0.88	0.96

* Estimates are rounded to the nearest hundred.

3. METHODS

Consider a simple ratio estimator

$$\hat{R} = \hat{Y} / \hat{X} ,$$

where \hat{Y} and \hat{X} denote estimators of two unknown population parameters Y and X . The Taylor series estimator of the variance of \hat{R} is

$$Var(\hat{R}) = \hat{R}^2 \left[\frac{Var(\hat{Y})}{\hat{Y}^2} + \frac{Var(\hat{X})}{\hat{X}^2} + 2 \frac{Cov(\hat{Y}, \hat{X})}{\hat{Y}\hat{X}} \right].$$

The estimated variance and covariance terms in the above formula should be specified with respect to both the sampling design and the form of the estimators \hat{Y} and \hat{X} . (For details, see Chapter 6 in Wolter, 1985).

Let \hat{q} be an estimator of a parameter of interest, q . And let $\hat{q}_{(a)}$ be the estimator of the same functional form as \hat{q} , but computed from the reduced sample obtained by omitting a -the group. The

jackknife estimator, $\hat{\mathbf{q}}$, is the mean of $\hat{\mathbf{q}}_{(a)}$,

$$\hat{\mathbf{q}} = \frac{\sum_{a=1}^k \hat{\mathbf{q}}_{(a)}}{k}.$$

The jackknife estimator of the variance is

$$\text{Var}\left(\hat{\mathbf{q}}\right) = \frac{1}{k(k-1)} \sum_{a=1}^k \left(\hat{\mathbf{q}}_a - \hat{\mathbf{q}}\right)^2.$$

For the simple ratio estimator, $\hat{R} = \hat{Y} / \hat{X}$,

$$\text{Var}\left(\hat{R}\right) = \frac{1}{k(k-1)} \sum_{a=1}^k \left(\hat{R}_a - \hat{R}\right)^2,$$

where $\hat{R} = \frac{\sum_{a=1}^k \hat{R}_{(a)}}{k}$. (For details, see Chapter 4 in Wolter, 1985).

The resampling variance estimates were computed in WesVar (Version 4.0) by specifying the “Jackknife 1” (JK1) option, which is one of five replication options, available in WesVar. Replication variance estimation involves calculating estimates for subsamples of the full sample and then computing the variance among the subsample estimates (replicates). The JK1 option should be applied when explicit stratification has not been used to select the sample, even though systematic sampling, which employs implicit stratification, has been used. Replicate weights can be imported into WesVar or can be created within the program by specifying the variance strata (VarStrat), the PSU identifier (VarUnit), and the replication method to be used.

Note that the replicate weights and the full sample weight must be specified for each analysis. Also, post-stratification, raking, and nonresponse adjustments can be applied to the replicate and full sample weights. However, the *Adult Caretaker Household Survey* full sample weight was adjusted for nonresponse and iterative poststratification (raking) was implemented before the full sample weight was incorporated into WesVar.

The Taylor series variance estimates were computed in SUDAAN (Version 6.4) by specifying the “with replacement” (WR) design option used for most clustered samples. Note that in the absence of design information (PSU and stratum identifiers), the WR option is often used to approximate variances for complicated designs (multi-stage, clustered, or stratified).

The SUDAAN sample design option determines the set of SUDAAN statements that are necessary for computing the variance of the specified design. The sample design statements that are required with the WR design option are the WEIGHT and NEST statements. The WEIGHT statement specifies the

variable whose values are the analysis weights or sample weights. The NEST statement specifies the variables whose values identify the PSUs and strata. The NISMART-2 sample design did not employ explicit stratification but implicit strata were formed (the 50 states and D.C.) during the systematic sample selection procedure. The WR option implies that the first stage sample units were sampled with replacement or with small sampling fractions in every first stage stratum. The first stage sampling units in the NISMART-2 design were households, sampled with equal probabilities of selection and *without* replacement. However, the “with replacement” design option was used since the first stage sampling fraction was small.

Due to the increase in the total number of children in the U.S. population under 18 years of age, from just over 63 million in 1988 to almost 72 million in 1999 (Bureau of the Census, 1988, 1999), the incidence estimates were standardized and reported as rates of the various types of missing children per 1,000 children age 0-17 in the population prior to conducting the tests for any change in the incidence rates.

4. RESULTS

The Taylor series and JK1 design effect estimates reported in Table 1 suggest that there will be some variation in the estimated variances, and these differences are reported in Table 2, which compares the Taylor series and JK1 standard errors and confidence interval limits. As indicated in Table 2, the difference in the standard error estimates for the incidence rates has a negligible effect on the lower and upper limits of the 95% confidence intervals around the estimates, and are expected to have no impact on the tests for change over time.

This expectation is supported by the t statistics and two-tailed p levels reported in Table 3, which provides the results of the tests for historical change in the incidence rates of missing children using the original NISMART-1 definitions. As indicated in Table 3, the test results yield the same conclusions for the significance of changes in the incidence rates of missing children between 1988 and 1999 regardless of whether the variances are computed with JK1 or Taylor series methods.

Table 2. Estimates of NISMART-1 Types of Missing Children Based on the 1999 NISMART-2 Data With Standard Errors and 95% Confidence Intervals

Type of NISMART-1 Episode	Rate per 1,000 Children Age 0-17	JK1 Standard Error	Taylor Series Standard Error	JK1 95% C.I.	Taylor Series 95% C.I.
Broad Scope Family Abduction	4.18	0.56	0.54	3.09-5.27	3.13-5.23
Policy Focal Family Abduction	3.15	0.44	0.45	2.28-4.02	2.27-4.04
Nonfamily Abduction	0.62	0.23	0.24	0.18-1.06	0.16-1.08
Broad Scope Runaway	5.28	0.43	0.51	4.43-6.14	4.28-6.29
Policy Focal Runaway	1.26	0.27	0.26	0.73-1.79	0.75-1.77
Broad Scope Lost, Injured, or Otherwise Missing	3.40	0.36	0.40	2.69-4.10	2.61-4.18
Policy Focal Lost, Injured, or Otherwise Missing	0.51	0.12	0.13	0.27-0.74	0.26-0.76

Table 3. Test for Change in the Incidence Rates of NISMART-1 Types of Missing Children 1988-1999 Based on JKI and Taylor Series Variance Estimates

Type of NISMART-1 Episode	Rate per 1,000 Children Age 0-17 1988	Rate per 1,000 Children Age 0-17 1999	Taylor Series t test Value	JK1 t test Value	Taylor Series p Value	JK1 p Value
Broad Scope Family Abduction	5.62	4.18	-1.81	-1.79	0.034360	0.032360
Policy Focal Family Abduction	2.59	3.15	0.87	0.88	0.522176	0.522176
Nonfamily Abduction	0.88	0.62	-0.45	-0.45	0.617077	0.610054
Broad Scope Runaway	7.09	5.28	-1.60	-1.64	0.064319	0.057439
Policy Focal Runaway	2.06	1.26	-1.34	-1.33	0.144296	0.147065
Broad Scope Lost, Injured, or Otherwise Missing	6.95	3.40	-3.17	-3.21	0.000809	0.000699
Policy Focal Lost, Injured, or Otherwise Missing	2.21	0.51	-1.83	-1.83	0.062891	0.062891

5. CONCLUSION

This paper compared the variances computed by jackknife 1 (JK1) and Taylor series methods using recently released data from a large-scale survey based on a national probability sample with a relatively “simple” complex sampling design of sufficient complexity to warrant the use of complex variance estimation methods. The results indicate that there were some small variations in the standard errors of the estimates computed with the two methods, however, these differences had negligible impact on the 95% confidence intervals and no effect on the conclusions drawn from statistical tests conducted to determine if there has been a change in the incidence rates of missing children in the U.S. over time.

REFERENCES

American Association for Public Opinion Research. (2000). *Standard Definitions: Final Disposition Codes and Outcome Rates for Surveys*. Ann Arbor, MI: AAPOR.

Department of Labor. (March, 1988) Current Population Survey. Washington: Bureau of the Census.

Department of Labor. (March, 1999) Current Population Survey. Washington: Bureau of the Census.

Hammer, H., Finkelhor, D.F., Sedlak, A.J., and Porcellini, L.E. (*Forthcoming*). *Historical Change in the Incidence of Missing, Abducted, Runaway, and Thrownaway Children, 1988-1999*. Washington, DC: Department of Justice, Office of Juvenile Justice and Delinquency Prevention.

Research Triangle Institute. (1995). SUDAAN™ 6.4. Research Triangle Park, NC: Research Triangle Institute.

Weng, S.S., Zhang, F, and Cohen, M.P. (1995). *Variance Estimates Comparison by Statistical Software*. ASA Proceedings of the Section on Survey Research Methods, pages 333-338. Arlington, VA: American Statistical Association.

Westat. (2000) WesVar™ 4.0. Westat, Inc.: Rockville, MD.

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