THE POWER TO KNOW.

SAS/STAT® 13.2 User's Guide

Introduction to Survey Sampling and Analysis Procedures

This document is an individual chapter from *SAS/STAT® 13.2 User's Guide*.

The correct bibliographic citation for the complete manual is as follows: SAS Institute Inc. 2014. *SAS/STAT® 13.2 User's Guide*. Cary, NC: SAS Institute Inc.

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513.

August 2014

SAS provides a complete selection of books and electronic products to help customers use SAS® software to its fullest potential. For more information about our offerings, visit **support.sas.com/bookstore** or call 1-800-727-3228.

Chapter 14
Introduction to Survey Sampling and Analysis Procedures

Contents

Overview: Survey Sampling and Analysis Procedures

This chapter introduces the SAS/STAT procedures for survey sampling and describes how you can use these procedures to analyze survey data.

Researchers often use sample survey methodology to obtain information about a large population by selecting and measuring a sample from that population. Because of variability among items, researchers apply probability-based scientific designs to select the sample. This reduces the risk of a distorted view of the population and enables statistically valid inferences to be made from the sample. For more information about statistical sampling and analysis of complex survey data, see Lohr (2010); Kalton (1983); Cochran (1977); Kish (1965). To select probability-based random samples from a study population, you can use the SURVEYSELECT procedure, which provides a variety of methods for probability sampling. To analyze sample survey data, you can use the SURVEYMEANS, SURVEYFREQ, SURVEYREG, SURVEYLOGISTIC, and SURVEYPHREG procedures, which incorporate the sample design into the analyses.

Many SAS/STAT procedures, such as the MEANS, FREQ, GLM, LOGISTIC, and PHREG procedures, can compute sample means, produce crosstabulation tables, and estimate regression relationships. However, in most of these procedures, statistical inference is based on the assumption that the sample is drawn from an infinite population by simple random sampling. If the sample is in fact selected from a finite population by using a complex survey design, these procedures generally do not calculate the estimates and their variances according to the design actually used. Using analyses that are not appropriate for your sample design can lead to incorrect statistical inferences.

The SURVEYMEANS, SURVEYFREQ, SURVEYREG, SURVEYLOGISTIC, and SURVEYPHREG procedures properly analyze complex survey data by taking into account the sample design. These procedures can be used for multistage or single-stage designs, with or without stratification, and with or without unequal weighting. The survey analysis procedures provide a choice of variance estimation methods, which include Taylor series linearization, balanced repeated replication (BRR), and the jackknife.

Table 14.1 briefly describes the SAS/STAT sampling and analysis procedures.

Table 14.1 Survey Sampling and Analysis Procedures in SAS/STAT Software

PROC SURVEYSELECT	
Selection Methods	Simple random sampling (without replacement) Unrestricted random sampling (with replacement) Systematic Sequential Bernoulli Poisson Probability proportional to size (PPS) sampling, with and without replacement PPS systematic PPS for two units per stratum PPS sequential with minimum replacement
Allocation Methods	Proportional Optimal Neyman
Sampling Tools	Stratified sampling Cluster sampling Replicated sampling Serpentine sorting

Table 14.1 *continued*

PROC SURVEYMEANS	
Statistics	Means and totals Proportions Quantiles Geometric means Ratios Standard errors Confidence limits
Analyses	Hypothesis tests Domain analysis Poststratification
Graphics	Histograms Box plots Summary panel plots Domain box plots
PROC SURVEYFREQ	
Tables	One-way frequency tables Two-way and multiway crosstabulation tables Estimates of population totals and proportions Standard errors Confidence limits
Analyses	Tests of goodness of fit Tests of independence Risks and risk differences Odds ratios and relative risks Kappa coefficients
Graphics	Weighted frequency and percent plots Mosaic plots Odds ratio, relative risk, and risk difference plots Kappa plots
PROC SURVEYREG	
Analyses	Linear regression model fitting Regression coefficients Covariance matrices Confidence limits Hypothesis tests Estimable functions Contrasts Least squares means (LS-means) of effects Custom hypothesis tests among LS-means Regression with constructed effects Predicted values and residuals Domain analysis
Graphics	Fit plots

Table 14.1 *continued*

PROC SURVEYLOGISTIC	
Analyses	Cumulative logit regression model fitting
	Logit, probit, and complementary log-log link functions
	Generalized logit regression model fitting
	Regression coefficients
	Covariance matrices
	Confidence limits
	Hypothesis tests
	Odds ratios
	Estimable functions
	Contrasts
	Least squares means (LS-means) of effects
	Custom hypothesis tests among LS-means
	Regression with constructed effects
	Model diagnostics
	Domain analysis
PROC SURVEYPHREG	
Analyses	Proportional hazards regression model fitting
	Breslow and Efron likelihoods
	Regression coefficients
	Covariance matrices
	Confidence limits
	Hypothesis tests
	Hazard ratios
	Contrasts
	Predicted values and standard errors
	Martingale, Schoenfeld, score, and deviance residuals
	Domain analysis

The Survey Procedures

The SURVEYSELECT procedure provides methods for probability sample selection. The SURVEYMEANS, SURVEYFREQ, SURVEYREG, SURVEYLOGISTIC, and SURVEYPHREG procedures provide statistical analyses for survey data. The following sections contain brief descriptions of these procedures. For more information, see the chapters for the individual procedures.

PROC SURVEYSELECT

The SURVEYSELECT procedure provides a variety of methods for selecting probability-based random samples. The procedure can select a simple random sample or can sample according to a complex multistage

sample design that includes stratification, clustering, and unequal probabilities of selection. With probability sampling, each unit in the survey population has a known, positive probability of selection. This property of probability sampling avoids selection bias and enables you to use statistical theory to make valid inferences from the sample to the survey population.

PROC SURVEYSELECT provides methods for both equal-probability sampling and probability proportional to size (PPS) sampling. Equal-probability sampling methods include simple random sampling (without replacement), unrestricted sampling (with replacement), systematic sampling, and sequential sampling.

In PPS sampling, a unit's selection probability is proportional to its size measure. PPS sampling is often used in cluster sampling, where you select clusters (groups of sampling units) of varying size in the first stage of selection. Available PPS methods include without replacement, with replacement, systematic, and sequential with minimum replacement. PROC SURVEYSELECT can apply these selection methods for stratified, clustered, and replicated sample designs.

For stratified sampling, PROC SURVEYSELECT provides survey design methods to allocate the total sample size among the strata. Available allocation methods include proportional, Neyman, and optimal allocation. Optimal allocation maximizes the estimation precision within the available resources, taking into account stratum sizes, costs, and variances.

For more information, see Chapter 102, "The SURVEYSELECT Procedure."

PROC SURVEYMEANS

The SURVEYMEANS procedure produces estimates of population means and totals from sample survey data. The procedure also computes estimates of proportions for categorical variables, estimates of quantiles for continuous variables, estimates of geometric means for positive continuous variables, and ratio estimates of means and proportions. For all these statistics, PROC SURVEYMEANS provides standard errors, confidence limits, and *t* tests when appropriate.

PROC SURVEYMEANS provides domain analysis, which computes estimates for domains (subpopulations), in addition to analysis for the entire study population. Formation of subpopulations can be unrelated to the sample design, and so the domain sample sizes can actually be random variables. Domain analysis takes this variability into account by using the entire sample to estimate the variance of domain estimates. Domain analysis is also known as subgroup analysis, subpopulation analysis, and subdomain analysis.

PROC SURVEYMEANS also performs poststratification, which adjusts the sampling weights so that their distribution matches known auxiliary information. Poststratification is often used to improve the efficiency of the analysis and adjust for nonresponse. PROC SURVEYMEANS provides poststratified analyses and also produces poststratified weights that can be used in the other survey analysis procedures. For more information about poststratification, see Fuller (2009); Lohr (2010); Wolter (2007); Rao, Yung, and Hidiroglou (2002).

PROC SURVEYMEANS uses ODS Graphics to create graphs as part of its output. Available statistical graphics include histograms and summary panel plots for continuous variables, box plots, and domain box plots.

For more information, see Chapter 99, "The SURVEYMEANS Procedure."

PROC SURVEYFREQ

The SURVEYFREQ procedure produces one-way to *n*-way frequency and crosstabulation tables from sample survey data. These tables include estimates of population totals, population proportions (overall proportions, and also row and column proportions), and corresponding standard errors. Confidence limits, coefficients of variation, and design effects are also available. The procedure provides the following types of design-adjusted confidence limits for proportions: Wald, logit, modified Wilson (score), and modified Clopper-Pearson (exact).

For one-way frequency tables, PROC SURVEYFREQ provides Rao-Scott chi-square goodness-of-fit tests, which are adjusted for the sample design. You can test a null hypothesis of equal proportions for a one-way frequency table, or you can input custom null hypothesis proportions for the test. For two-way frequency tables, PROC SURVEYFREQ provides design-adjusted tests of independence, or no association, between the row and column variables. These tests include the Rao-Scott chi-square test, the Rao-Scott likelihood ratio test, the Wald chi-square test, and the Wald log-linear chi-square test.

For 2×2 tables, PROC SURVEYFREQ computes estimates and confidence limits for risks (or row proportions), the risk difference, the odds ratio, and relative risks. For square tables, PROC SURVEYFREQ computes simple and weighted kappa coefficients.

PROC SURVEYFREQ uses ODS Graphics to create graphs as part of its output. Available statistical graphics include weighted frequency and percent plots, which can be displayed as bar charts or dot plots in various formats. For two-way tables, PROC SURVEYFREQ also produces mosaic plots. For multiway tables, PROC SURVEYFREQ also produces odds ratio, relative risk, risk difference, and kappa coefficient plots.

For more information, see Chapter 97, "The SURVEYFREQ Procedure."

PROC SURVEYREG

The SURVEYREG procedure performs regression analysis for sample survey data. The procedure fits linear models and computes regression coefficients and their variance-covariance matrices. PROC SURVEYREG enables you to specify classification effects by using the same syntax that the GLM procedure uses.

PROC SURVEYREG provides hypothesis tests for the model effects. The procedure also provides custom hypothesis tests for linear combinations of the regression parameters. The procedure computes confidence limits for the parameter estimates and also for any specified linear functions of the regression parameters. The procedure can produce an output data set that contains the predicted values from the linear regression, their standard errors and confidence limits, and the residuals.

PROC SURVEYREG also performs regression analysis for domains.

PROC SURVEYREG uses ODS Graphics to create graphs as part of its output. For models that depend on at most one regressor excluding the intercept, the procedure produces fit plots, which can be displayed as bubble plots or heat maps. In bubble plots, the bubble area is proportional to the observation's weight. In heat maps, the heat color represents the sum of the weights at the corresponding location.

For more information, see Chapter 101, "The SURVEYREG Procedure."

PROC SURVEYLOGISTIC

The SURVEYLOGISTIC procedure provides logistic regression analysis for sample survey data. Logistic regression analysis investigates the relationship between discrete responses and a set of explanatory variables. PROC SURVEYLOGISTIC fits linear logistic regression models for discrete response survey data by the method of maximum likelihood and incorporates the sample design into the analysis. The SURVEYLOGISTIC procedure enables you to specify categorical classification variables (also known as CLASS variables) as explanatory variables in the model by using the same syntax for main effects and interactions as in the GLM and LOGISTIC procedures.

The following link functions are available for regression in PROC SURVEYLOGISTIC: the cumulative logit function (CLOGIT), the generalized logit function (GLOGIT), the probit function (PROBIT), and the complementary log-log function (CLOGLOG). The procedure performs maximum likelihood estimation of the regression coefficients with either the Fisher scoring algorithm or the Newton-Raphson algorithm.

PROC SURVEYLOGISTIC also performs logistic regression analysis for domains.

For more information, see Chapter 98, "The SURVEYLOGISTIC Procedure."

PROC SURVEYPHREG

The SURVEYPHREG procedure performs regression analysis based on the Cox proportional hazards model for sample survey data. Cox's semiparametric model is widely used in the analysis of survival data to estimate hazard rates when explanatory variables are available. The regression coefficients are estimated by maximizing a pseudo-partial-likelihood function that incorporates the sampling weights. The procedure provides design-based variance estimates, confidence intervals, and tests for the estimated regression coefficients.

PROC SURVEYPHREG provides hypothesis tests for the model effects. The procedure also provides custom hypothesis tests for linear combinations of the regression parameters. The procedure computes hazard ratios and their confidence limits. The procedure can produce several observation-level output statistics, such as predicted values and their standard errors, martingale residuals, Schoenfeld residuals, score residuals, and deviance residuals.

PROC SURVEYPHREG also performs proportional hazards regressions for domains.

For more information, see Chapter 100, "The SURVEYPHREG Procedure."

Survey Design Specification

Survey sampling is the process of selecting a probability-based sample from a finite population according to a sample design. You then collect data from these selected units and use them to estimate characteristics of the entire population.

A *sample design* encompasses the rules and operations by which you select sampling units from the population and the computation of sample statistics, which are estimates of the population values of interest. The objective of your survey often determines appropriate sample designs and valid data collection methodology. A complex sample design can include stratification, clustering, multiple stages of selection, and unequal

weighting. The survey procedures can be used for single-stage designs or for multistage designs, with or without stratification, and with or without unequal weighting.

To analyze your survey data with the SURVEYMEANS, SURVEYFREQ, SURVEYREG, SURVEYLOGISTIC, and SURVEYPHREG procedures, you need to specify sample design information for the procedures. This information can include design strata, clusters, and sampling weights. All the survey analysis procedures use the same syntax for specifying sample design information. You provide sample design information with the STRATA, CLUSTER, and WEIGHT statements, and with the RATE= or TOTAL= option in the PROC statement.

If you provide replicate weights for BRR or jackknife variance estimation, you do not need to specify a STRATA or CLUSTER statement. Otherwise, you should specify STRATA and CLUSTER statements whenever your design includes stratification and clustering.

When there are clusters (PSUs) in the sample design, the procedures estimate variance by using the PSUs, as described in the section "Variance Estimation" on page 248. For a multistage sample design, the procedures use only the first stage of the sample design for variance estimation. Therefore, the required input includes only first-stage cluster (PSU) and first-stage stratum identification. You do not need to input design information about any additional stages of sampling.

The following sections provide brief descriptions of basic sample design concepts and terminology used in the survey procedures. For more information, see Lohr (2010); Kalton (1983); Cochran (1977); Kish (1965).

Population

Population refers to the target population, which is the group of units (individuals or elements) of interest for study. Often, the primary objective is to estimate certain characteristics of this population, which are called *population values*. A *sampling unit* is an individual or element in the target population. A *sample* is a subset of the population that is selected for the study.

Before you use the survey procedures, you should have a well-defined target population, sampling units, and an appropriate sample design.

In order to select a sample according to your sample design, you need to have a list of sampling units in the population. This is called a *sampling frame*. PROC SURVEYSELECT uses probability-based selection methods to select a sample from a sampling frame.

Stratification

Stratified sampling involves selecting samples independently within strata, which are nonoverlapping subgroups of the survey population. Stratification controls the distribution of the sample size in the strata. It is widely used to meet a variety of survey objectives. For example, with stratification you can ensure adequate sample sizes for subgroups of interest, including small subgroups, or you can use stratification to improve the precision of overall estimates. To improve precision, units within strata should be as homogeneous as possible for the characteristics of interest.

Clustering

Cluster sampling involves selecting clusters, which are groups of sampling units. For example, clusters might be schools, hospitals, or geographical areas, and sampling units might be students, patients, or citizens. Cluster sampling can provide efficiency in frame construction and other survey operations. However, it can also result in a loss in precision of your estimates, compared to a nonclustered sample of the same size. To minimize this effect, units within clusters should be as heterogeneous as possible for the characteristics of interest.

Multistage Sampling

In *multistage sampling*, you select an initial (first-stage) sample that is based on groups of elements in the population, which are called *primary sampling units* (*PSUs*).

Then you create a second-stage sample by drawing a subsample from each selected PSU in the first-stage sample. By repeating this operation, you can select a higher-stage sample. If you include all the elements from the selected primary sampling units, then the two-stage sample is a cluster sample.

Sampling Weights

Sampling weights, which are also known as *survey weights*, are positive values associated with the units in your sample. Ideally, the weight of a sampling unit should be the "frequency" that the sampling unit represents in the target population.

Often, sampling weights are the reciprocals of the selection probabilities for the sampling units. When you use PROC SURVEYSELECT, the procedure generates the sampling weight component for each stage of the design, and you can multiply these sampling weight components to obtain the final sampling weights. Sometimes, sampling weights also include nonresponse adjustments, poststratification, or regression adjustments by using supplemental information.

When the sampling units have unequal weights, you must provide the weights to the survey analysis procedures. If you do not specify sampling weights, the procedures use equal weights in the analyses.

Population Totals and Sampling Rates

If you use Taylor series variance estimation, the survey procedures include a finite population correction factor in the analysis if you input either the sampling rate or the population total.

The sampling rate is the ratio of the sample size (the number of sampling units in the sample) n to the population size (the total number of sampling units in the target population) N, $f = n/N$. This ratio is also called the *sampling fraction*. If you select a sample without replacement, the extra efficiency compared to selecting a sample with replacement can be measured by the *finite population correction* (*fpc*) factor, (1-f).

To include a finite population correction factor in your analysis, you can input either the sampling rate or the population total. Otherwise, the procedures do not use the *fpc* in computing variance estimates. For

fairly small sampling fractions, it is appropriate to ignore this correction. For more information, see Cochran (1977) and Kish (1965).

As discussed in the section "Variance Estimation" on page 248, for a multistage sample design, the procedures use only the first stage of the sample design for variance estimation. Therefore, if you are specifying the sampling rate, you should input the *first-stage sampling rate*, which is the ratio of the number of PSUs in the sample to the total number of PSUs in the target population.

If you use BRR or jackknife variance estimate, the procedures do not include a finite population correction in the analysis, and you do not need to input the sampling rate or the population total.

Variance Estimation

The survey analysis procedures provide a choice of variance estimation methods for complex survey designs. In addition to the Taylor series linearization method, the procedures offer two replication-based (resampling) methods—balanced repeated replication (BRR) and the delete-1 jackknife. These variance estimation methods usually give similar, satisfactory results (Lohr 2010; Särndal, Swensson, and Wretman 1992; Wolter 2007). The choice of a variance estimation method can depend on the sample design used, the sample design information available, the parameters to be estimated, and computational issues. For more information, see Lohr (2010).

The Taylor series linearization method is appropriate for all designs where the first-stage sample is selected with replacement, or where the first-stage sampling fraction is small, as it often is in practice. The Taylor series method obtains a linear approximation for the estimator and then uses the variance estimate for this approximation to estimate the variance of the estimate itself (Fuller 1975; Woodruff 1971). When there are clusters (PSUs) in the sample design, the procedures estimate the variance from the variation among the PSUs. When the design is stratified, the procedures pool stratum variance estimates to compute the overall variance estimate.

For a multistage sample design, the Taylor series method uses only the first stage of the sample design. Therefore, the required input includes only first-stage cluster (PSU) and first-stage stratum identification. You do not need to input design information about any additional stages of sampling.

Replication methods for variance estimation draw multiple replicates (or subsamples) from the full sample by following a specific resampling scheme. Commonly used resampling schemes include *balanced repeated replication* (BRR) and the *jackknife*. The parameter of interest is estimated from each replicate, and the variability among the replicate estimates is used to estimate the overall variance of the parameter estimate.

The BRR variance estimation method requires a stratified sample design with two PSUs in each stratum. Each replicate is obtained by deleting one PSU per stratum according to the corresponding Hadamard matrix and adjusting the original weights for the remaining PSUs. The adjusted weights are called *replicate weights*. The survey procedures also provide Fay's method, which is a modification of the BRR method.

The jackknife method deletes one PSU at a time from the full sample to create replicates, and modifies the original weights to obtain replicate weights. The total number of replicates equals the number of PSUs. If the sample design is stratified, each stratum must contain at least two PSUs, and the jackknife is applied separately within each stratum.

Instead of having the survey procedures generate replicate weights for the analysis, you can directly input your own replicate weights. This can be useful if you need to do multiple analyses with the same set of replicate weights, or if you have access to replicate weights without complete design information.

See the chapters on the survey procedures for complete details. For more information about variance estimation for sample survey data, see Lohr (2010); Wolter (2007); Särndal, Swensson, and Wretman (1992); Lee, Forthofer, and Lorimor (1989); Cochran (1977); Kish (1965); Hansen, Hurwitz, and Madow (1953).

Example: Survey Sampling and Analysis Procedures

This section demonstrates how you can use the survey procedures to select a probability-based sample and then analyze the survey data to make inferences about the population. The analyses include descriptive statistics and regression analysis. This example is a survey of income and expenditures for a group of households in North Carolina and South Carolina. The goals of the survey are as follows:

- Estimate total income and total living expenses
- Estimate the median income and the median living expenses
- Investigate the linear relationship between income and living expenses

Sample Selection

To select a sample with PROC SURVEYSELECT, you input a SAS data set that contains the sampling frame (the list of units from which the sample is to be selected). You also specify the selection method, the desired sample size or sampling rate, and other selection parameters. PROC SURVEYSELECT selects the sample and produces an output data set that contains the selected units, their selection probabilities, and their sampling weights. For more information, see Chapter 102, "The SURVEYSELECT Procedure."

In this example, the sample design is a stratified sample design, with households as the sampling units and selection by simple random sampling. The SAS data set HHFrame contains the sampling frame, which is the list of households in the survey population. The sampling frame is stratified by the variables State and Region. Within strata, households are selected by simple random sampling. The following PROC SURVEYSELECT statements select a probability sample of households according to this sample design:

```
proc surveyselect data=HHFrame out=HHSample
                  method=srs n=(3, 5, 3, 6, 2);
   strata State Region;
run;
```

The STRATA statement names the stratification variables State and Region. In the PROC SURVEYSELECT statement, the DATA= option names the SAS data set HHFrame as the input data set (or sampling frame) from which to select the sample. The OUT= option stores the sample in the SAS data set named HHSample. The METHOD=SRS option specifies simple random sampling as the sample selection method. The N= option specifies the stratum sample sizes.

The SURVEYSELECT procedure then selects a stratified random sample of households and produces the output data set HHSample, which contains the selected households together with their selection probabilities and sampling weights. The data set HHSample also contains the sampling unit identification variable Id and the stratification variables State and Region from the input data set HHFrame.

Survey Data Analysis

You can use the SURVEYMEANS and SURVEYREG procedures to estimate population values and perform regression analyses for survey data. The following example briefly shows the capabilities of these procedures. For more information, see Chapter 99, "The SURVEYMEANS Procedure," and Chapter 101, "The SURVEYREG Procedure."

The following PROC SURVEYMEANS statements estimate the total income and living expenses for the survey population based on the data from the stratified sample design:

```
proc surveymeans data=HHSample sum median;
   var Income Expense;
   strata State Region;
   weight Weight;
run;
```

The PROC SURVEYMEANS statement invokes the procedure, and the DATA= option names the SAS data set HHSample as the input data set to be analyzed. The keywords SUM and MEDIAN request estimates of population totals and medians.

The VAR statement specifies the two analysis variables Income and Expense. The STRATA statement names the stratification variables State and Region. The WEIGHT statement specifies the sampling weight variable Weight.

You can use PROC SURVEYREG to perform regression analysis for survey data. Suppose that, in order to explore the relationship between household income and living expenses in the survey population, you choose the following linear model:

$$\text{Expense} = \alpha + \beta * \text{Income} + \text{error}$$

The following PROC SURVEYREG statements fit this linear model for the survey population based on the data from the stratified sample design:

```
proc surveyreg data=HHSample;
   strata State Region ;
   model  Expense = Income;
   weight Weight;
run;
```

The STRATA statement names the stratification variables State and Region. The MODEL statement specifies the model, with Expense as the dependent variable and Income as the independent variable. The WEIGHT statement specifies the sampling weight variable Weight.

References

Cochran, W. G. (1977), *Sampling Techniques*, 3rd Edition, New York: John Wiley & Sons.

Fuller, W. A. (1975), "Regression Analysis for Sample Survey," *Sankhyā, Series C*, 37, 117–132.

Fuller, W. A. (2009), *Sampling Statistics*, Hoboken, NJ: John Wiley & Sons.

Hansen, M. H., Hurwitz, W. N., and Madow, W. G. (1953), *Sample Survey Methods and Theory*, volume 1 and 2, New York: John Wiley & Sons.

Kalton, G. (1983), *Introduction to Survey Sampling*, Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-035, Beverly Hills, CA: Sage Publications.

Kish, L. (1965), *Survey Sampling*, New York: John Wiley & Sons.

Lee, E. S., Forthofer, R. N., and Lorimor, R. J. (1989), *Analyzing Complex Survey Data*, Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-071, Beverly Hills, CA: Sage Publications.

Lohr, S. L. (2010), *Sampling: Design and Analysis*, 2nd Edition, Boston: Brooks/Cole.

Rao, J. N. K., Yung, W., and Hidiroglou, M. A. (2002), "Estimating Equations for the Analysis of Survey Data Using Poststratification Information," *Sankhyā*, 64, Series A, 364–378.

Särndal, C. E., Swensson, B., and Wretman, J. (1992), *Model Assisted Survey Sampling*, New York: Springer-Verlag.

Wolter, K. M. (2007), *Introduction to Variance Estimation*, 2nd Edition, New York: Springer.

Woodruff, R. S. (1971), "A Simple Method for Approximating the Variance of a Complicated Estimate," *Journal of the American Statistical Association*, 66, 411–414.

Index