

SURV 699E / SURVMETH 895

Practical Tools for Sampling and Weighting

Instructors: Richard Valliant

Meeting Time and Places: Thursday 3:30 - 6:10 pm
2208 Lefrak Hall and 4036 ISR

Office Hours: By appointment

Office Phone: 301-405-0932

Email: rvalliant@survey.umd.edu

Class Website: www.jpsm.umd.edu/surv699e/
Class notes and datasets for homework will be on the website,
which is accessible from outside the UMD network.
Username/Password: will be distributed by email

Video site: <http://www.jpsmclasses.umd.edu/Mediasite/Viewer/>
All lectures will be recorded and available on the mediasite 1 or 2
days after the class.

Instructions for remote access to JPSM network for running SAS at home:

www.jpsm.umd.edu/support.htm

For help contact dgilbert@survey.umd.edu

Home SAS licenses: <http://www.oacs.umd.edu/software/softdis.asp>.

SAS documentation: <http://support.sas.com/onlinedoc/913/docMainpage.jsp>
http://support.sas.com/documentation/onlinedoc/91pdf/index_913.html

R Downloadable for free from <http://cran.r-project.org/>

Grades:	Homeworks	20%
	3 Projects	25% each
	Class participation	5%

Synopsis

This course will be a combination of hands-on applications and general review of the theory and methods behind different approaches to sampling and weighting. Several survey data sets will be used to illustrate how to design samples, calculate weights, and make estimates from complex surveys. For statisticians, the course is meant to give you some practical experience in applying

the theoretical ideas you have learned in this and other classes. For social scientists, the goal is to provide you some insight into the statistical thinking needed and steps taken in actually designing, selecting, and weighting samples. By working together in teams, you should get a taste of how projects are carried out in survey organizations.

Project 1. The first project will be to design a sample for a population where single-stage sampling is appropriate. A stratified sample will be designed that achieves specified goals for precision of a series of domain estimates while meeting a budget constraint. Anticipated rates of nonresponse and ineligible units will be accounted for. Multicriteria optimization methods will be used to determine the allocation.


Project 2. The second project will be to use a set of data collected from a sample of military personnel and develop survey weights. The weights should account for cases with unknown eligibility, nonrespondents, and uses of auxiliary data to improve estimators. Based on a description of the survey design along with variables that identify cases as respondents and nonrespondents, teams will compute weights, document their methods, discuss pros and cons of different options, and make class presentations on the results. Students will devise quality control checks and will set up the analysis file to allow use of either linearization or replication variance estimation.

Project 3. A third application will be an area probability design in which students will use an existing sample of primary units and determine a plan for sampling segments and persons within segments. Rates will be determined to achieve target sample sizes for different demographic groups.

Readings for the course will cover the general theory and application of alternative methods of sample allocation, nonresponse adjustment (including weighting cell adjustments and propensity score adjustments), methods of using auxiliary data to reduce coverage biases and improve standard errors (including poststratification and more general applications of regression weighting). Readings will also include alternatives that are used for computing response and

coverage rates. Throughout the course the emphasis will be on learning how to apply the methods rather than on learning the theory behind them.

Tentative Class Schedule¹

Class	Date	Description
1	4 Sep 08	<p>1. Sample Design and Sample Size</p> <ul style="list-style-type: none"> - Choosing a sample design - Goals of a design - Frames vs. target populations - Determining the sample size for a single-stage design - Domain estimates - Estimating population parameters from a sample <p>Readings/References</p> <p>Särndal, Swensson, and Wretman (1992), <i>Model Assisted Survey Sampling</i>, chapters 3, 12.</p> <p>Victor, R.G., et al. (2004), “The Dallas Heart Study: A Population-Based Probability Sample for the Multidisciplinary Study of Ethnic Differences in Cardiovascular Health,” <i>The American Journal of Cardiology</i>, 93, 1473-1480.</p> <p>Project 1 is assigned</p>
2	11 Sep 08	<p>Finish Ch. 1</p> <p>Review of  software</p> <ul style="list-style-type: none"> - Installation - Importing data - Editing functions, scripts with RWinEdt - Vector-matrix computations
3	18 Sep 08	<p>2. Proposals</p> <ul style="list-style-type: none"> - Overview of proposal process - Budgets - Proposal Writing <p>3. Study Performance rates</p> <ul style="list-style-type: none"> - Examples of disposition codes - Rate definitions (response, contact, location, etc.) - Inflating sample sizes for loss of analysis cases - Weighted vs. unweighted rates

¹ Updates to the class schedule will be posted to the Web site as needed.

		<p>Readings/References</p> <p>The American Association for Public Opinion Research (2006), Standard Definitions Final Dispositions of Case Codes and Outcome Rates for Surveys.</p> <p>Abraham, K., Maitland, A., and Bianchi, S. (2006), “Nonresponse in the American Time Use Survey Who Is Missing from The Data and How Much Does It Matter?” <i>Public Opinion Quarterly</i>, 70, 676-703.</p> <p>Callegaro, M. et al. (2007), “Fitting disposition codes to mobile phone surveys: experiences from studies in Finland, Slovenia and the USA,” <i>Journal of the Royal Statistical Society A</i>, 170, 647-670.</p>
4	25 Sep 08	<p>4. Power Computations</p> <ul style="list-style-type: none"> - One-sample and two-sample tests - One-sided and two-sided tests - Tests for continuous variables - Tests for proportions <p>Readings/References</p> <p>Cohen, J. (1992), “A Power Primer,” <i>Psychological Bulletin</i>, 112, 155-159.</p> <p>Black, K. (2006). R Tutorial. http://www.cyclismo.org/tutorial/R/index.html.</p> <p>Park, H.M. (2004). “Understanding the Statistical Power of a Test.” http://www.indiana.edu/~statmath/stat/all/power/power.pdf</p>
5	2 Oct 08	<p>5. Mathematical Programming</p> <ul style="list-style-type: none"> - Mathematical formulation; Multiple criteria - Excel Solver - SAS proc nlp <p>Proposal for Project 1 is due</p>
6	9 Oct 08	<p>6. Steps in Weighting</p> <ul style="list-style-type: none"> - Base weights – various designs - Adjustments of unknown eligibility - Nonresponse adjustments <ul style="list-style-type: none"> - cell adjustments - propensity score adjustments - Other adjustments <ul style="list-style-type: none"> - Coverage - What to do with ineligible - Use of auxiliary data <ul style="list-style-type: none"> - poststratification - general regression estimation - calibration estimation

		<ul style="list-style-type: none"> - Weight variability - Variance inflation / unequal weighting effects - Weight trimming - Constrained adjustments - Software <ul style="list-style-type: none"> - Cell formation using classification algorithms: RPART, CHAID, CART - Propensity adjustments - IVEware <p>Readings/References</p> <p>Kalton, G., and Flores-Cervantes, I. (2003), "Weighting Methods," <i>Journal of Official Statistics</i>, 19, 81-97.</p> <p>Battaglia, et al. (1995), Adjusting for Noncoverage of Nontelephone Households in the National Immunization Survey, <i>Proceedings of the Section on Survey Methods Research</i>, American Statistical Association, 678-683.</p> <p>Potter, F. (1993), "Effect of Weight Trimming," <i>Proceedings of the Section on Survey Methods Research</i>, American Statistical Association, 758-763.</p> <p>Hoaglin, D., and Battaglia, M. (1996), "A Comparison of Two Methods of Adjusting for Noncoverage of Nontelephone Households in a Telephone Survey," <i>Proceedings of the Section on Survey Methods Research</i>, American Statistical Association, 497-502.</p> <p>Project 2 is assigned</p>
7	16 Oct 08	<p>Project 1—Class presentations</p> <p>Project 1 reports are due on 14 Oct 08</p>
8	23 Oct 08	<p>7. Quality Control</p> <ul style="list-style-type: none"> - Specification writing - Flowcharts - Quality control of weighting steps - Archiving programs, data sets
9	30 Oct 08	<p>8. Theory for Weight Adjustments</p> <ul style="list-style-type: none"> - Nonresponse - Coverage - Totals vs. means; model parameters - Auxiliary data—use of models to decide <p>Readings/References</p> <p>Little, R. (1986), "Survey Nonresponse Adjustments for Estimates of</p>

		<p>Means,” <i>International Statistical Review</i>, 54, 139-157.</p> <p>Little, R., and Vartivarian, S. (2003), “On weighting the rates in non-response weights,” <i>Statistics in Medicine</i>, 22, 1589-1599.</p> <p>Little, R., and Vartivarian, S. (2003), “Does Weighting for Nonresponse Increase the Variance of Survey Means?,” <i>Survey Methodology</i>, 31, 161-168.</p> <p>Spencer, B. (2000), “___”, <i>Survey Methodology</i>, 28, ___.</p>
10	6 Nov 08	<p>9. Variance Estimation</p> <ul style="list-style-type: none"> - Exact methods - Linearization - Replication—random groups, jackknife, BRR - Effects of different steps in weighting on variance estimates <ul style="list-style-type: none"> - How to account for with linearization - Repeating adjustments in each replicate - FPCs - Software <p>Readings/References</p> <p>Rust, K., and Rao, J.N.K. (1996), “Variance Estimation for Complex Survey Using Replication Techniques,” <i>Statistical Methods in Medical Research</i>, 5, 283-310.</p> <p>Rust, et al. (2006), “Finite Population Correction Factors (Panel Discussion),” <i>Proceedings of the Section on Survey Methods Research</i>, American Statistical Association.</p>
11	13 Nov 08	<p>10. Area Sampling</p> <ul style="list-style-type: none"> - Stages of sampling - Sources of data - Geographic breakdowns, census data available - Supplemental information—address lists (Genesys, InfoUSA) <ul style="list-style-type: none"> - Relisting—half-open interval method to picked up omissions <p>Readings/References</p> <p>Morton, et al. (2006), “2005 National Survey on Drug Use and Health,” prepared by Research Triangle Institute for the Substance Abuse and Mental Health Services Administration.</p> <p>National Center for Health Statistics (1992), <i>Sample Design: Third National Health and Nutrition Examination Survey</i>, Series 2: Data Evaluation and Methods Research, No. 113, U.S. Dept. of Health and Human Services.</p> <p>Project 3 is assigned</p>

12	20 Nov 08	Project 2—Class presentations Project 2 reports due
	27 Nov 08	No class—Thanksgiving
13	4 Dec 08	11. Designing PSU Samples <ul style="list-style-type: none"> - National survey examples, CPS, NSDUH, CBECS - Variance components <ul style="list-style-type: none"> - General component formulas <ul style="list-style-type: none"> - Different numbers of stages of sampling - Intraclass correlations - How many PSUs? - Sizes of PSU variance components in national surveys - Estimation of components
14	11 Dec 08	Project 3—Class presentations Project 3 reports due