The Analysis of Survey Data
Political Science 519 / SOC 519 – Fall 2018
Mondays & Wednesdays 7:00 – 8:30 PM in Willard 174

Professor: Dr. Eric Plutzer (EXP12, 865-6576)
Office Hours: *By appointment* Monday & Wednesdays 9:45-11:00 in Pond 312

Teaching Assistant: Mr. Erik Hernandez (ELH209)
Office Hours: Mondays 11:30am – noon, and 1pm to 2pm in Oswald Tower 318

Social scientists use surveys to address important theoretical puzzles and to inform public policy. However, survey data rarely satisfies the definition of a simple random sample. If the survey's response rate is not 100%, then the sample is not a simple random sample. If the survey design over-represented some groups deliberately, or if interviews were conducted in selected neighborhoods or other clusters, then we need to account for that. Missing data due to unanswered questions also poses a challenge to valid statistical analysis.

In all likelihood, your previous statistics classes introduced you to techniques that assume a simple random sample. Applying these methods to survey data will frequently yield results that are **wrong**. Fortunately, survey statisticians have developed methods that address all of these challenges. You will learn about these methods in PLSC/SOC 519.

This focus is statistical, but with an emphasis on the intuition and logic underlying the statistical models rather than focusing on proofs and estimation. This will provide a foundation for frequent hands-on applications in this class and provide a formal understanding that will prepare students to tackle intermediate and advanced topics independently or in advanced courses.

This class builds on your prior coursework and experience:

This is an *intermediate* level course in quantitative analysis for students who have completed 2 semesters of graduate-level statistics (not general research methods or mixed methods) and who are interested in the application of social statistics to data collected by surveys. This course will build on your prior knowledge of techniques such as analysis of variance, ordinary least squares regression, and measures of uncertainty (e.g., confidence intervals and standard errors). Students who are shaky on these three building blocks have done well in this class, but only when they take steps to review these fundamentals and work especially hard to keep up.

All students should have experience writing *replicable data management and data analysis scripts* in Stata (do files), SPSS (syntax files), SAS or R. We will use the Stata statistical package and experience with other packages will easily transfer for those new to Stata.

Learning Objectives:

Students will learn how decisions concerning research design have specific and well-understood consequences for common statistical methods.

Students will learn how to create and apply design weights and post-stratification weights.

Students will learn the underlying theory and statistical logic that provides the foundation of techniques to calculate unbiased measures of uncertainty (standard errors and confidence intervals) for common survey designs.

Students will learn the principles of missing data imputation and learn how to employ current best practices when using survey data sets with missing data.

Students will learn one or more additional special topics useful in survey analysis. The topics vary from year to year, and may include contextual analysis and geocoding, measurement and psychometrics, machine coding of open-ended questions, analysis of repeated cross-sectional surveys, and the design and analysis of non-probability online panels.

Evaluation:

15% of the course grade will be based on homework. Homework, assigned 15-20 times during the semester, is due before class begins.   Students earn 9 points out of 10 with demonstrated effort (you complete the assignment), timely submission, and (when relevant) the integration of theory with applied work. If you simply "don't get it," please do your best and we will discuss an opportunity to resubmit the assignment. Homework assignments that strike me as "exceptional" in terms of clarity, thoroughness, insightfulness or creativity may receive a grade of 10. Homework that appears rushed, incomplete, or unclear will receive lower scores.

Each student may submit two assignments late (submitted in class when we next meet) without penalty regardless of reason. But do your best to save these for times when you really need them (e.g., illness, professional travel, etc.). Other late assignments lose one point each class session.

> I will attempt to assign less challenging homeworks to be due on Wednesdays, but this may not always be possible.

5% is based on your attendance, preparation to contribute to class discussions, and participation. I expect you to: (a) read and think about assigned readings before coming to class, (b) think critically about your own homework. Do not fall into the trap of leaving the written homework until late the night before class as this will make it difficult to get through the assigned reading.

50% will be based on two in-class pencil and paper exams (25% each) that focus on sampling, weighting, and the analysis of complex surveys.

30% of the course grade will be based on a final project analyzing data sets that you select. The final paper must incorporate the following techniques: (a) analysis of weighted data, (b) analysis of clustered data, (c) missing data imputation, and (d) one additional topic covered in class. Not all data sets will permit all four kinds of analysis. As a result, multiple data sets may be required.

Required textbooks:

Groves, Fowler, Couper, Lepkowski, Singer & Tourangeau. *Survey Methodology.* Wiley. 1st or 2nd edition is acceptable.

For reference, any textbook (including online) that covers difference of means, difference of proportions, correlation, regression, analysis of variance, and (for each of these) the calculation of confidence intervals and significance tests.

Book available at no cost via libraries.psu.edu

Allison. 2002. "Missing Data." *Quantitative Applications in the Social Sciences* #136 (Sage).

TENTATIVE SCHEDULE AND READING ASSIGNMENTS:

Introduction and Review

Monday, August 20, 2018. Introduction to the course: Overview and introduction to telephone poll methodology.

Wednesday, August 22, 2018. Introduction to secondary analysis of survey data sets.

> **Read:**
>> Heering, West & Berglund. Chapter 1 in *Applied Survey Data Analysis.* [Canvas]
>>
>> Groves et al. Chapter 1

Monday, August 27, 2018. Introduction to Survey Analysis; statistics review

> Complete: The anonymous class survey at this link: https://pennstate.qualtrics.com/jfe/form/SV_3CqxXLKlgZmaHJj
>
> Read:, and sections 10.1 – 10.4
>
> Review formulae for basic descriptive statistics and confidence intervals for univariate statistics.
>
> ***Students who are new to Stata***: Now is a good time to familiarize yourself with Stata. There are many online resources, but here are two to check out:
>
>> http://data.princeton.edu/stata/ and http://www.ats.ucla.edu/stat/stata/modules/

Wednesday, August 29, 2018.   Secondary Data Sets

> Read: General Social Survey Codebook, Appendix A (2006 edition). Take your time, and read every word, making a list of key concepts and jargon.
>
> Read: The sampling design section from a codebook for any standard study used in your field of study. This might be a study whose data you may wish to use for some of the assignments in this class.
>
> Come to class prepared to discuss this data set. Use section 1.3 in Groves as a guide for things to think about. Ask me for suggestions if you are unfamiliar with such studies.

Wednesday, September 5, 2018. Review of Complex Survey Design Basics

> Read Groves et al. Chapters 2-3 and section 4.3.

<u>Review</u> the logic of the "finite correction factor" that probably was discussed in your elementary statistics textbook (typically, when covering variance estimation for samples drawn with and without replacement).

Part I: Introduction to sampling statistics

Monday, September 10, 2017. Complex Phone Surveys; Introduction to Stata

<u>Read</u>: Voss, Gelman and King. 1995. "A review: Preelection survey methodology: Details from eight polling organizations, 1988 and 1992." *Public Opinion Quarterly* 59 (Spring): 98-132. [JSTOR]

<u>Read</u>: Ralph Frerichs, Equal Probability of Selection (posted to CANVAS)

Wednesday, September 12, 2018. Unequal selection of primary sampling units.

<u>Read</u>: Groves et al. chapter 5 and 10.5.

Monday, September 17, 2018.   Weighting for design choices that result in unequal probabilities

<u>Read</u>: Groves et al. 10.5.1 – 10.5.2

Wednesday, September 19, 2018. Weighting adjustments for unit non-response.

<u>Read</u>: Groves et al. 10.5.3 and 10.5.4

Monday, September 24, 2018. Post-stratification weighting to match known population totals.

<u>Read</u>: Bataglia et al. 2006. "Weighting Survey Data." AAPOR Annual Meeting. [CANVAS]

<u>Optional</u>: Kolinikov, Stas. 2014. "Calibrating survey data using iterativeproportional fitting (raking)." *Stata Journal.*

Wednesday, September 26, 2018. **Generating post-stratification weights by ratios and raking**

Groves et al 10.5.5

Winship and Radbill (1994) "Sampling Weights and Regression Analysis." *Sociological Methods Research* vol 23: 230-257. [CANVAS]

Monday, October 1, 2018. Analyzing weighted data

Wednesday, October 3, 2018. ** MIDTERM #1 *****

Fundamentals of sampling and inference:
Confidence interval of means & proportions
SRS, EPSEM, PPS, and Mitovsky-Waksburg Sampling

Ratio weights, raking, and analysis of weighted data using `[aw=` and `[pw=`

Monday, October 8, 2018. Introduction to the analysis of cluster and complex surveys.

Groves et al., Chapter 4 and supplemental online materials, TBA

Stata manual sections on svy routines to be assigned

Wednesday, October 10, 2018. Design effects and effective sample size.

Groves et al. 10.7

Monday, October 15, 2018. Software and practical considerations with complex surveys having complete design information

Eugene Johnson. 1989. "Considerations and Techniques for the Analysis of NAEP Data." *Journal of Educational Statistics* 14 (4): 303-334. [Posted to CANVAS].

David Johnson and Lisa Elliott. 1998. "Sampling Design Effects: Do They Effect Analyses from the Nation Survey of Families and Households?" *Journal of Marriage and the Family* 60 (Nov): 993-1001.

Wednesday, October 17, 2018. Software and practical considerations with complex surveys having replicate weights.

Wesvar Manual, Appendix A, pages A1 – A31. [Posted to CANVAS]

 "Replicate Weights." UCLA: Statistical Consulting Group. from http://www.ats.ucla.edu/stat/stata/library/replicate_weights.htm/   (accessed October 7, 2015).

Part II: Topics and skills in the survey analyst's toolkit

Monday, October 22, 2018. Introduction to Missing Data

Allison. 2002. "Missing Data." *QASS* (Sage Little Green Books), pp 1-27.

Stata MI Reference Manual, as assigned in homework

Wednesday, October 24, 2018. Missing Data and Multiple Imputation

Allison. 2002. "Missing Data." *QASS* (Sage Little Green Books), pp 27-50.

Stata MI Reference Manual, as assigned in homework

Monday, October 29, 2018. Multiple Imputation in Stata; MI in practice

Johnson, D. R. and Young, R. (2011), Toward Best Practices in Analyzing Datasets with Missing Data: Comparisons and Recommendations. *Journal of Marriage and Family*, 73: 926–945. 5

Wednesday, October 31, 2018. Detailed discussion of final project expectations, and Introduction to Matching Methods for Non-Probability Sampling.

Ansolabehere, S., & Rivers, D. (2013). Cooperative survey research. *Annual Review of Political Science*, *16*, 307-329.

Rivers, D., & Bailey, D. (2009). Inference from matched samples in the 2008 US national elections. In *Proceedings of the Joint Statistical Meetings* (pp. 627-639).

Optional: Baker, Reg, J. Michael Brick, Nancy A. Bates, Mike Battaglia, Mick P. Couper, Jill A. Dever, Krista J. Gile, and Roger Tourangeau. "Report of the AAPOR task force on non-probability sampling."A (May, 2013).

Monday, November 5, 2018   *** Midterm Exam #2 ***

Weighting (carryover),	Stratification
FPC	Clustering
Design effect	Use of svyset
Multiple Imputation	

Wednesday, November 7, 2018. Measurement – a formal look at reliability and validity

[I know this is a lot of reading immediately after the midterm, but give yourself some time to read through once]

Read: Bohrnstedt, "Measurement models for survey research." Pp. 347-404 in Marsden, P. V., & Wright, J. D. (2010). Handbook of survey research (2nd ed.). Bingley, UK: Emerald.

Read: Stata 14 Manual sections on: `factor` and `alpha`

Monday, November 12, 2018 – Introduction to contextual analysis

Read 3 of the following papers or any three papers drawn from these or from your discipline that use contextual measures as independent variables to predict individual behavior:

Banaszak, Lee Ann and Eric Plutzer. 1993. "Contextual Determinants of Feminist Attitudes: National and Subnational Influences in Western Europe." *American Political Science Review* 87 (March):147-57.

Ken Goldstein & Paul Freedman (2000) "New Evidence for New Arguments: Money and Advertising in the 1996 Senate Elections." *Journal of Politics* 62: 1087-1108.

Marylee C. Taylor (1998) "How White Attitudes Vary with the Racial Composition of Local Populations: Numbers Count." *American Sociological Review* 63: 512-535.

Huckfeldt, Robert, Eric Plutzer, & John Sprague. 1993. "The Alternative Contexts of Political Behavior: Churches, Neighborhoods, and Individuals." *Journal of Politics* 55 (May):365-81.

***   Preliminary outline for final project due (ungraded, but must be submitted) ***

Wednesday, November 14, 2018. –   Adding Environmental Context measures from secondary data sets to Survey Data II.

Krieger N, Waterman PD, Chen JT, Rehkopf DH, & Subramanian SV. "Geocoding and monitoring US socioeconomic inequalities in health: an introduction to using area-based socioeconomic measures -- The Public Health Disparities Geocoding Project monograph." Boston, MA: Harvard School of Public Health. Available at: http://www.hsph.harvard.edu/thegeocodingproject/

Read sections:

Executive summary: https://www.hsph.harvard.edu/thegeocodingproject/executive-summary/

Introduction: https://www.hsph.harvard.edu/thegeocodingproject/introduction/

Geocoding: https://www.hsph.harvard.edu/the**geocoding**project/**geocoding**

Generating ABSMs: https://www.hsph.harvard.edu/thegeocodingproject/generating-absms/

Case example: https://www.hsph.harvard.edu/thegeocodingproject/case-example/

or

https://cdn1.sph.harvard.edu/wp-content/uploads/sites/2068/2017/03/The-Public-Health-Geocoding-Project-Case-Example-03_07_17.pdf

Skim sections:

Analytic methods: https://www.hsph.harvard.edu/thegeocodingproject/analytic-methods/

Multilevel modeling: https://www.hsph.harvard.edu/thegeocodingproject/multi-level-modeling/

Visual display: https://www.hsph.harvard.edu/thegeocodingproject/visual-display/

Monday, November 19, 2018 & Wednesday, November 21, 2018 – No Class: Thanksgiving Holiday

Monday, November 26, 2018. Creating Environmental Context measures from your primary survey.

> Bradford S. Jones & Barbara Norrander. (1996) "The Reliability of Aggregated Public Opinion Measures." *American Journal of Political Science* 40: 295-309.

> Optional: O'Brien, Robert M. (1990) "Estimating the reliability of aggregate-level variables based on individual-level characteristics." *Sociological Methods & Research* 18: 473-504.

Wednesday, November 28, 2018. Creating Environmental Context measures from your primary survey (continued).

> TBA

Monday, December 3, 2018.   Review and catchup

Wednesday, December 5, 2018, Review and catchup

Monday, December 10, 2018. Final Project due before midnight.

Additional disclosures and policies

Please read each of the following sections carefully. You are responsible for being familiar with all relevant policies, even though it may be tempting to skip the "fine print."

Academic Integrity Policy

The Department of Political Science, along with the College of the Liberal Arts and the University, takes violations of academic dishonesty seriously. Observing basic honesty in one's work, words, ideas, and actions is a principle to which all members of the community are required to subscribe.

You must work independently on the final project, with the exception that you may seek assistance from colleagues or classmates for narrow, technical issues such as identifying bugs in a statistics script (do-file). You are required to acknowledge this assistance in a footnote that describes the problem, the assistance you received, and the names of all who provided assistance. You may not have assistance in the interpretation of statistical output or in explaining the methods you employ.

Your final project can have shared content with other papers you are writing for another class, but only with written permission, in advance, from me and the other professor.

Collaboration on homework is permitted, provided you include a detailed description of the collaboration/assistance. You are not permitted to complete homework through a division of labor (each person answers a different question) and you must be the author of all assignments submitted under your name.

To earn a passing grade, you must properly cite sources that form the basis for your review of literature, development of hypotheses and models, descriptions of data, results and interpretations. References should be as specific as possible (page numbers, chapter numbers, etc.) and allow any reader to easily find the material and locate the specific basis of your claims. Direct quotations, paraphrases and key ideas should all be properly attributed.

Before each examination, you will receive instructions on any notes or "cheat sheets" that are permitted, as well as the exams calculator policy. Other kinds of assistance are not permitted. Both the giving and the receipt of other kinds of assistance is a violation of academic integrity and will be reported to the College and/or Graduate School. Lying to the instructor or purposely misleading any Penn State administrator shall also constitute a violation of academic integrity. Federal law prohibits tampering with, disabling, or destroying lavatory smoke detectors (this last sentence is known as an *attention check*).

In cases of any violation of academic integrity, it is the policy of the Department of Political Science to follow procedures established by the College of the Liberal Arts.

Note to students with disabilities:

Penn State welcomes students with disabilities into the University's educational programs. If you have a disability-related need for reasonable academic adjustments in this course, contact the Office for Disability Services (ODS) at 814-863-1807 (V/TTY). For further information regarding ODS, please visit the Office for Disability Services Web site at http://equity.psu.edu/ods/.

In order to receive consideration for course accommodations, you must contact ODS and provide documentation (see the documentation guidelines at http://equity.psu.edu/ods/guidelines/documentation-guidelines).

If the documentation supports the need for academic adjustments, ODS will provide a letter identifying appropriate academic adjustments. Please share this letter and discuss the adjustments with your instructor as early in the course as possible. You must contact ODS and request academic adjustment letters at the beginning of each semester.

Extended Absences:

During your enrollment at Penn State, unforeseen challenges may arise. If you ever need to miss an extended amount of class in such a circumstance, please notify me so you can determine the best course of action to make up missed work. If your situation rises to a level of difficulty you cannot manage on your own with faculty support, reach out to the Student Care & Advocacy office by phone at (814-863-2020) or email them at StudentCare@psu.edu. Office hours are Monday-Friday, 8 a.m. to 5 p.m.

Privacy and self-disclosure

In this class, you may be asked to play the role of a survey respondent. When doing so, you need not give truthful answers about yourself, your opinions or experiences because this is a role-playing exercise. Likewise, when a classmate plays the role of a survey respondent, you should assume that answers about their health, traumatic experiences, and values are fictititious. Disclosing personal information will never be a required element of the course.

Sexual Harassment & Assault

Federal law, Title IX, and Penn State policy prohibits discrimination, harassment, and violence based on sex and gender (including sexual harassment, sexual assault, domestic/dating violence, stalking, sexual exploitation and retaliation). If you or someone you know has been harassed or assaulted, you can receive confidential counseling support through PSU's Gender Equity Center (204 Boucke, 863-2027) or the Centre County Women's Resource Center's 24/7 hotline (877-234-5050). Alleged violations can be reported non-confidentially to the PSU Title IX Coordinator (867-0099, titleix@psu.edu). Reports to law enforcement can be made to PSU's Police Department (863-1111).

I will seek to keep information you share with me private to the greatest extent possible, but *as a professor I am designated as a "responsible employee" and have mandatory reporting responsibilities* to share information when I learn about crimes or sexual misconduct. This policy may make it difficult for faculty to provide as much one-on-one support as they might prefer, but is intended to help make our campus a safer place for all.