Political Science 3802: Survey Research and Design

MW 3:30-5pm

LSRM 112B (3231 Walnut St.)

Fall 2022

Course Instructor

Dr. William Marble
Office: 33 Fox-Fels Hall (3814 Walnut St., 3rd Floor)
Office Hours: Wednesday 10am–12pm, or by appointment
Email: marblew@sas.upenn.edu

Overview and Goals

In modern society, surveys are everywhere. Surveys are used to estimate the unemployment rate, to assess public opinion about current events such as the repeal of *Roe v. Wade*, to predict which party will control the House and Senate after this year's midterms, to generate television ratings, and much more. As these examples suggest, surveys are one of the predominant tools to understand politics, the economy, consumer behavior, and more. From universities to corporate headquarters, newsrooms to Capitol Hill, survey research can be found in every sector in society.

This class will teach students how to design, conduct, and analyze surveys, as well as to communicate the results of surveys to a broad audience. Students will learn how to be a critical consumer of information derived from surveys. To gain practical experience, the course will provide many opportunities to work hands-on with real political polling data collected ahead of the 2022 midterm elections.

This class will begin with an overview of the role of surveys in contemporary society. We will then cover principles of survey design, including defining the research question and how to write effective questionnaires to answer the research question.

Next, we will discuss some basic probability theory and statistical sampling. We'll learn an "idealized" version of sampling and methods for analyzing "ideal" survey data. But as recent high-profile polling errors have demonstrated, real-world surveys often fail to meet the "ideal" standards. For example, fewer than 5% of people who are asked to take a survey will do so, and those who agree to participate may be very different from those who do not. How do departures from the idealized survey collection affect our conclusions? And (how) can we correct for these errors? In answering these questions — which are the subject of active scientific and industry research — students will get hands-on experience applying state-of-the-art statistical methods to survey data.

Finally, we will introduce a number of advanced topics which may include: forecasting and poll aggregation; survey experiments; measuring ideology and other latent traits; panel surveys; and methods to reduce measurement error.

Much of the material will be technical in nature, but just as important is the ability to interpret the results of surveys substantively, relate them to overarching questions of interest, and explain the methodology (including strengths and limitations) to non-experts. As such, course assignments will cover both technical and writing/communication skills, and course readings will be a mix of technical material and applied survey research.

Prerequisites

PSCI 3800 (formerly 107) or equivalent approved by instructor. You should have experience performing the following tasks in the R programming language: reading in data, subsetting data, generating new variables, merging datasets, and calculating descriptive statistics such as means and standard deviations. Prior knowledge of inferential statistics (e.g., measuring uncertainty, hypothesis testing, and linear regression) will be helpful but is not required.

Computation

Survey research is an inherently quantitative field that requires data manipulation and statistical analysis. Throughout the course, we will weave together theory with computational tools for data manipulation and statistical analysis. We will work in the open-source statistical programming language R, which is used extensively in both academia and industry settings.

You should download and install R and RStudio on your computer. R is the underlying programming language, while RStudio is a program that makes it easier to use R. You can download R from r-project.org and RStudio from rstudio.com/products/rstudio/.

Assignments

Paper Discussion. On select class meetings, three students will be assigned to lead a discussion of one of the supplemental readings assigned for the week. The group leading the discussion should come prepared to summarize the paper's objective, methods, and results, and should have several follow-up questions prepared to kick off the discussion. Each group will present twice throughout the semester. More instructions will be provided in class.

Problem Sets. There will be 4 problem sets throughout the course of the semester. The problem sets will be a mix of writing and data analysis, which will require coding in R. The problem sets are meant to give you a chance to build real-world skills in designing and analyzing surveys. I encourage you to work on the problem set in groups, but each student must submit their own write-up. If you work with other students on the problem set, please note who you worked with in your submission.

Midterm Exam. There will be one closed-book midterm exam which will will focus on conceptual understanding and ability to communicate the topics of the course. While this may involve some explanation of technical material, the exam will not focus on mathematical formulas or computer coding. It will be given in class on **Monday, October 24**.

Final Research Project. For your final project, you will be tasked with proposing a research question that is interesting to you, finding and analyzing survey data to answer that question, and to write up the results. You have two options on what data to use:

1. Our class is taking place during a midterm election year, which provides a convenient opportunity to use real political survey data. In early November — around the time of the election — I will provide a survey dataset that you can analyze to answer a question of your choosing related to voting in the 2022 midterm elections.
2. You can find your own survey dataset. More detailed suggestions on how to find survey data will be distributed later in the semester.

Your write-up should include a clearly stated research question, a brief discussion of why that question is important, a description of the dataset you used to answer the question, a methodology section that explains how you analyzed the data, a results section that presents your answer to the question (with tables and

graphs where appropriate), and a discussion section where you can comment on the strength of the evidence you obtained and suggest follow-up research. While there is no formal length requirement, I expect that it will take around 10 pages to cover these requirements (double-spaced, including tables and figures but excluding references). Additionally, you will submit the data and code that you used to produce your analysis.

Finally, on the last day of class, students will present their projects and get feedback from the class which they can incorporate into the final paper. The presentations do not need to be too detailed, but they should cover the main research question, the data/methods used to answer the question, and present any preliminary results.

Grading

Attendance, participation, engagement, and office hours attendance: 15%

- This is a small upper-level seminar, and it is crucial that students attend class, participate, and engage with peers.
- Class will be structured so that there are many ways to participate: by asking questions, by participating in small group discussions, and so on.
- You are also required to come to office hours at least once before the end of October. This will be an opportunity for you to ask questions, discuss ideas for your final project, or talk about other ideas and interests. Additionally, it helps me to understand what I need to do to make sure everyone in the class can succeed.

Group paper discussion: 15%

- Groups will be graded on the quality of the discussion that they lead, not on their mastery of the material they present.
- For example, a high-quality discussion could focus on resolving confusion about a paper; it could focus on critiquing the paper's methodology; or it could focus on follow-up questions that the paper raises.

Problem sets: 30%

- Scored on a 1 to 12 scale. Getting all the questions "correct" will translate into a score of 10.
- Scores of 11 and 12 will be reserved for submissions that have all the correct answers, have written explanations that convey an excellent understanding of the course material, and (where applicable) have code and analysis that is particularly well-executed and clearly communicated.
- Students that average 10 our of 12 on problem sets should expect a final grade in at least the B+/A- range.
- No penalty for your first late submission, as long as it is turned in within 3 days of the due date. Any late submissions after that will receive a zero, unless you have a valid (university designated) excuse.

Midterm exam: 15%

- This will be an in-class midterm on **Monday, October 24**.

Final project: 25%

- 5% for your presentation on the last day of class
- 20% for the final submission

Textbook and Other Reading

The main textbook will be:

> Groves, Robert M., Floyd J. Fowler Jr., Mick P. Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau. 2009. *Survey Methodology*, 2nd edition. Wiley.

This textbook is available online, for free, through the Penn Library.

There are additional required readings, which are typically academic research articles but also include some newspaper articles and blog posts. Each topic also has a number of supplemental readings. I encourage you to read at least one of the supplemental readings for each topic — even when you are not assigned to present one of them. All additional readings will be posted on Canvas.

You should come prepared to discuss the reading on the day it is assigned. E.g., by 9/7 you should have looked at Groves chs. 1-2.

A note on how to read research papers. Being able to quickly read and digest information is an important skill that takes active development. For this class, I don't expect you to read every word in the assigned reading, nor are you expected to understand every bit of what you do read. Instead, you should read enough to be able to summarize the main points of the paper, the supporting evidence, and understand how the paper relates to the class topics. Typically, this means carefully reading the abstract and introduction, then *purposively skimming* the rest of the paper. By *purposively skimming*, I mean that you should read all of the section headings, and then pick out sections of the paper that you need to read more closely in order to answer the following questions: What question or problem does this paper address? What makes this problem interesting or important? What methods does the paper propose to use to address the problem, and why? What is the paper's answer to the question or solution to the problem? How does this paper relate to other research? For a useful guide on how researchers tend to read papers, see "How to Read a Paper" by S. Keshav: https://web.stanford.edu/class/ee384m/Handouts/HowtoReadPaper.pdf

Finally, from time to time you will read something and find yourself confused. **That is okay!** It happens to everyone. It might feel discouraging at first, but it's also an opportunity to learn. When you find that you don't understand a paper, the first thing to do is to try your best to answer the questions above — preferably in writing. Then, write down any further questions or areas of confusion, and bring them to class or office hours so that we can discuss them. If you have a question, it's very likely that someone else in the class has the same or a similar question.

Office Hours and Contact Information

I will hold office hours on Wednesday mornings from 10am to noon in my office in Fox-Fels Hall, room 33 (enter on Walnut Street and walk up the stairs to the third floor). If you cannot meet at that time, please email me and we can find another time to meet. I am also happy to chat directly before or after class.
I try to respond to email within 24 hours, though I may be slower on weekends. If you do not get a reply from me within two days, feel free to send me a follow-up email. Many issues are better answered in person, so I encourage you to talk to me directly for questions beyond simple course logistics.

Academic Integrity

I expect all students to abide by the rules of the University and to abide by the University's Code of Academic Integrity.[1] You are expected to do your own work and to cite sources, ideas, and words that you

[1] `https://catalog.upenn.edu/pennbook/code-of-academic-integrity/`

borrow from others.[2]

Collaboration is an important part of survey research, and I encourage you to work on your assignments with your classmates. Everything you turn in, however, should be your writing and/or code. You should write at the top of the assignment which classmates you worked with. Additionally, re-using an assignment from another class is not permitted, unless you have explicit permission from both me and the instructor of the other class.

There is one important caveat about plagiarism norms in survey research. It is standard practice to copy existing survey questions verbatim from other sources — especially "gold standard" surveys like the American National Election Studies or the General Social Survey. This is good practice to enable comparison across different surveys. While you don't need to formally cite the source of a question, you should indicate in brackets where you got the survey question (e.g., [ANES 2020]) to help me understand how you drafted your questionnaire.

[2]The exact format of the citation is unimportant, so long as the citation contains the author, title, and any other information necessary for me to find the original source online or in the library.

Course Schedule

N.B.: This schedule represents my current plans. As the semester continues, we may need to adjust the schedule. These changes will be announced on Canvas and the syllabus will be updated accordingly.

8/31 - Introduction and Overview of Class

Required Reading

- The course syllabus

9/7 - Bird's-Eye View of Survey Research

Note: There is no class 9/5 due to Labor Day

Topics

- Examples of government, academic, media, and industry surveys; uses of surveys; overview of survey research process; sources of error in surveys

Required Reading

- Groves, ch. 1. Skim all the example surveys, but pick one to read in depth.
- Groves, ch. 2. Focus especially on the "Quality Perspective" perspective in §2.3.

Supplemental Reading

- Venkatesh Rao. 2010. "A Big Little Idea Called Legibility." Ribbonfarm blog. [Link]
- Daniel M. Butler and David W. Nickerson. 2011. "Can Learning Constituency Opinion Affect How Legislators Vote? Results from a Field Experiment." *Quarterly Journal of Political Science* 6(1): 55-83. [Canvas]
- Lee, Melissa M., and Nan Zhang. 2017. "Legibility and the Informational Foundations of State Capacity." *Journal of Politics* 79(1): 118-132. [Canvas]

9/12 and 9/14 - Data Manipulation and Analysis in R

Topics

- Variable types; loading data; recoding variables and generating new variables; subsetting; summary statistics; merging datasets; for loops; sampling and simulation

Preparation

- Make sure you have R and RStudio installed. Detailed instructions are available on Canvas.
- If you have trouble, email me or come to office hours.
- Bring your laptop to class.

9/19 and 9/21 - Probability Theory and Sampling Theory

Problem Set 1 Distributed 9/21

Topics

- Defining the population; random sampling; sampling methods; mean, variance, and correlation; basics of weighting

Required Reading

- Groves ch. 3
- Squire, Peverill. 1988. "Why the 1936 Literary Digest Poll Failed." *Public Opinion Quarterly* 52(1): 125–33. [Canvas]
- Ansolabehere, Stephen, and Brian F. Schaffner. 2014. "Does Survey Mode Still Matter? Findings from a 2010 Multi-Mode Comparison." *Political Analysis* 22(3): 285–303. [Canvas]

9/26 and 9/28 - Questionnaire Design

Topics

- Validity and reliability; Cognitive processes of survey-taking; order and priming effects; interviewer effects; mode effects; attention checks

Required Reading

- Groves chs. 7-8

Supplemental Reading

- Krosnick, Jon A. 1991. "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys." *Applied Cognitive Psychology* 5: 213–236. [Canvas]

10/3 and 10/5 - Nonresponse: Causes, Consequences, and Mitigation

Problem Set 1 Due at the Beginning of Class 10/3
Problem Set 2 Distributed 10/5

Topics

- Unit and item nonresponse; survey weighting; problems of selection on unobservables; social trust; using administrative data to estimate response probabilities

Required Reading

- Groves ch. 6 and ch. 10.5-10.6
- Barreto, Matt, Chris Warshaw, Matthew A. Baum, Bryce J. Dietrich, Rebecca Goldstein, and Maya Sen. April 22, 2019. "New research shows just how badly a citizenship question would hurt the 2020 Census." *The Washington Post.* [Link]
- Podcast: "How Much Are Polls Misrepresenting Americans?" on *The Science of Politics*, hosted by Matt Grossman, featuring interviews of Josh Clinton and Amnon Cavari. [Link]

Supplemental Reading

- Clinton, Joshua D., John S. Lapinski, and Marc J. Trussler. 2022. "Reluctant Republicans, Eager Democrats? Partisan Nonresponse and the Accuracy of 2020 Presidential Pre-Election Telephone Polls." *Public Opinion Quarterly* (Forthcoming). [Canvas]
- Cavari, Amnon, and Guy Freedman. 2022. "Survey Nonresponse and Mass Polarization: The Consequences of Declining Contact and Cooperation Rates." *American Political Science Review.* [Canvas]
  - And a critique: Mellon, Jonathan, and Christopher Prosser. 2022. "Correlation with Time Explains the Relationship between Survey Nonresponse and Mass Polarization." *Journal of Politics.* [Canvas]

10/10 and 10/12 - Horserace Polling and Election Forecasting

Topics

- Likely voter models; differential nonresponse; poll aggregation; generic ballot polling; issue knowledge; prediction markets

Required Reading

- AAPOR Ad Hoc Committee on 2016 Election Polling. "An Evaluation of 2016 Election Polls in the U.S." [Link]
- Keeter, Scott, and Ruth Igielnik. 2016. "Can Likely Voter Models Be Improved? Evidence from the 2014 U.S. House elections." Pew Research Center. Sections 2 ("Measuring the likelihood to vote") and Section 3 ("Comparing the results of different likely voter models"). [Link]
- Silver, Nate, and FiveThirtyEight staff. 2022. "How FiveThirtyEight's House, Senate And Governor Models Work." [Link]

Supplemental Reading

- Gelman, Andrew, and Gary King. 1993. "Why Are American Presidential Election Campaign Polls So Variable When Votes Are So Predictable?" *British Journal of Political Science* 23(4): 409–51. [Canvas]
- Rogers, Todd, and Masahiko Aida. 2014. "Vote Self-Prediction Hardly Predicts Who Will Vote, and Is (Misleadingly) Unbiased." *American Politics Research* 42(3): 503–28. [Canvas]

10/17 and 10/19 - Estimating Public Opinion in Small Subgroups or Geographic Areas

Problem Set 1 Due at the Beginning of Class 10/3

Topics

- Targeted samples; oversamples; stratified sampling; combining surveys; (multilevel) regression and poststratification

Required Reading

- Groves TBD
- Ghitza, Yair, and Jonathan Robinson. 2020. "What Happened in 2020." Catalist blog post. [Link]

Supplemental Reading

- Fraga, Bernard L., Yamil R. Velez, and Emily A. West. 2022. "Reversion to the Mean, or Their Version of the Dream? An Analysis of Latino Voting in 2020." Working paper. [Canvas]
- Lax, Jeffrey R., and Justin H. Phillips. 2009. "Gay Rights in the States: Public Opinion and Policy Responsiveness." *American Political Science Review* 103(03): 367–386. [Canvas]

10/24 and 10/26 - Measuring Ideology and Other Latent Traits

Problem Set 2 Due at the Beginning of Class 10/24

Topics

- Conceptualizing latent traits; direct elicitation; developing scales; measuring validity and reliability of scales; item-response theory models

Required Reading

- Ellis, Christopher, and James Stimson. 2012. *Ideology in America*. Cambridge University Press. Excerpts. [Canvas]
- Klein, Ezra. 2015. "No one's less moderate than moderates." Vox.com. [Link]
- Jefferson, Hakeem. 2022. "The Curious Case of Black Conservatives: Construct Validity and the 7-point Liberal-Conservative Scale." [Canvas]

Supplemental Reading

- Gerber, Elizabeth R., and Jeffrey B. Lewis. 2004. "Beyond the Median: Voter Preferences, District Heterogeneity, and Political Representation." *Journal of Political Economy* 112(6): 1364–1383. [Canvas]
- Hill, Seth J., and Chris Tausanovitch. 2015. "A Disconnect in Representation? Comparison of Trends in Congressional and Public Polarization." *Journal of Politics* 77(4): 1058–75. [Canvas]
- Ahler, Douglas J., and David E. Broockman. 2018. "The Delegate Paradox: Why Polarized Politicians Can Represent Citizens Best." *Journal of Politics* 80(4): 1117–33. [Canvas]
- Foley, Patrick, and John McDonnell. 2017. "What the SATs Taught Us about Finding the Perfect Fit." Stitch Fix blog (really!). [Link]

10/31 - In-Class Midterm Exam

- Exam will cover topics through 10/19

11/2 - Polling on the 2022 Midterm Elections

Topics

- Hands-on work with pre-election survey data

Required Reading

- TBD: light reading on 2022 midterm polling

11/7 and 11/9 - Survey Experiments and Causal Inference

Problem Set 3 Distributed 11/9

Topics

- Overview of causal inference; split ballot designs; vignette experiments; conjoint experiments

Required Reading

- TBD
- Grady, Christopher. "10 Things to Know About Survey Experiments." EGAP. [Link]

Supplemental Reading

- Hainmueller, Jens, and Daniel J. Hopkins. 2015. "The Hidden American Immigration Consensus: A Conjoint Analysis of Attitudes toward Immigrants." *American Journal of Political Science* 59(3): 529–48. [Canvas]
- Jensen, Amalie, William Marble, Kenneth Scheve, and Matthew J. Slaughter. 2020. "City Limits to Partisan Polarization in the American Public." *Political Science Research and Methods* 9(2): 223–41. [Canvas]

11/14 and 11/16 - Social Desirability Bias and Measurement of Sensitive Topics

Topics

- Principles of social desirability bias; turnout over-reporting; list experiments; randomized response designs; mode differences; implicit association tests

Required Reading

- Groves TBD
- Coppock, Alexander. October 25, 2016. "Shy Trump supporters? This new evidence says no." *The Washington Post.* [Link]
- Nosek, Brian A., Anthony G. Greenwald, and Mahzarin R. Banaji. 2007. "The Implicit Association Test at Age 7: A Methodological and Conceptual Review." In *Automatic Processes in Social Thinking and Behavior*, J.A. Bargh (ed.). Psychology Press. [Canvas]

Supplemental Reading

- Lyall, Jason, Graeme Blair, and Kosuke Imai. 2013. "Explaining Support for Combatants during Wartime: A Survey Experiment in Afghanistan." *American Political Science Review* 107(4): 679–705. [Canvas]
- Tourangeau, Roger, and Ting Yan. 2007. "Sensitive Questions in Suveys." *Psychological Bulletin* 133(5): 859-883. [Canvas]
- Jee, Haemin, and Tongtong Zhang. 2021. "Opposing Autocracy without Support for Democracy: A Study of Non-democratic Critics in China." [Canvas]
- Iyengar, Shanto, and Sean J. Westwood. 2015. "Fear and Loathing across Party Lines: New Evidence on Group Polarization." *American Journal of Political Science* 59(3): 690–707. [Canvas]

11/21 and 11/28 - Panel Surveys and Time Series Cross-Sectional Surveys

Note: Class on Wednesday, 11/23 is canceled due to Thanksgiving
Problem Set 3 Due at Beginning of Class 11/21
Problem Set 4 Distributed 11/21

Topics

- Rolling cross-sections; repeated interviewing; panel attrition; stability of attitudes; causal inference with panel data; difference-in-differences; interrupted time series designs; within estimators

Required Reading

- Groves
- Converse, Philip E. 1964. "The Nature of Belief Systems in Mass Publics." In *Ideology and Discontent*, ed. David Apter. New York: The Free Press pp. 206–261. [Canvas]
- Card, David, and Alan B. Krueger. 1994. "Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania." *American Economic Review* 84(4): 772-793. [Canvas]

Supplemental Reading

- Margalit, Yotam. 2013. "Explaining Social Policy Preferences: Evidence from the Great Recession." *American Political Science Review* 107(1): 80–103. [Canvas]
- Broockman, David, and Joshua Kalla. 2016. "Durably reducing transphobia: A field experiment on door-to-door canvassing." *Science* 352(6282): 220-224. [Canvas]

11/30 - 2022 Midterm Elections Redux

Topics

- Assessment of election polling; analysis of trends in voting blocs

Required Reading

- TBD: News analysis of 2022 midterm polls
- Grimmer, Justin, William Marble, and Cole Tanigawa-Lau. 2022. "Measuring the Contribution of Voting Blocs to Election Outcomes." [Canvas]

12/5 - Additional Topics TBD

- Potential topics include: inattentive survey respondents; exit polling; combining survey data with aggregate data; incentivized survey games; working with Census data; surveying hard-to-reach populations; working with open-ended responses;
- Feel free to suggest topics!

12/12 - Research Presentations

12/22 - Final Paper Due

- Email by the end of the day

Syllabus last updated on August 23, 2022