Statistics 4750
Sample Survey Methods
Syllabus
Fall 2024

Professor: Dr. Elaine Zanutto
E-mail: zanutto@wharton.upenn.edu
ezanutto@gmail.com

Office Hours: Wednesdays at 6pm via Zoom

Teaching Assistant: TBA
TA Office hours: TBA
Class Website: Canvas website

Course Description: This course will focus on statistical methods for the design and analysis of sample surveys. We plan to cover Chapter 1-5 and parts of Chapters 6-9, and 11 of the textbook, and other related topics as time permits. The core topics are
- simple random sampling
- stratified sampling
- cluster sampling
- ratio estimation
- sampling with unequal probabilities
- sampling weights
- analyzing complex surveys
- dealing with nonresponse
- some advice for questionnaire design
- pros and cons of online panels
- linear and logistic regression with survey data
- graphing survey data with R
- analysis of categorical data from a complex survey
- small area estimation

Course Requirements: There will be four homework assignments, a project (including an oral presentation), and a midterm exam.

Homework Policy: Homeworks must be turned in individually. You may speak to each other about the homework. You may ask me and our TA about the homework. But you must turn in your own work. The midterm exam will be closed-book but a formula sheet will be provided for you.
Late Policy for Homework: Extensions to homework deadlines are often granted. Homeworks are due on Wednesday nights. Generally anyone who asks can have an extension until Saturday midnight. Send me an email to ask for an extension and cc our TA.

Missed Homework and Missed Midterm Policy: Missing homeworks are not easily accommodated. Accommodation will be made without penalty ONLY in the event of an excused absence including serious illness. Note that homework extensions are often granted so please take advantage of that option. Missing the midterm is extremely difficult to accommodate and will require documentation for absence to avoid a grade of zero. Specific accommodations for missing work are granted on a case-by-case basis and may include make-up work or shifting grade weight to the midterm exam or final project. Any work missing without permission may be given a grade of zero.

Grading:

Homework 30% (all 4 are weighted equally)
Midterm 40% (closed-book, formula sheet will be provided for you)
Project Essay 10% (individual project, not a group project)
Project presentation 10% (powerpoint presentation in class or pre-recorded)
Project question/answer period 10% (in-class)

The midterm will be closed-book but I will provide a formula sheet.

Attendance: Attendance at lecture is not mandatory. I will record all lectures and post them on the canvas website (though occasionally technical glitches do occur, in which case you should get notes from a fellow classmate). However, it is your responsibility to watch the lectures. Do not come to office hours unprepared.

Computing software: We will make use of the statistical computing software R via the RStudio interface in class and on homework assignments. R can be downloaded from http://www.r-project.org/ and RStudio can be downloaded from https://posit.co/products/open-source/rstudio/. It would be beneficial if you already have some experience with R, but not absolutely necessary. I will give you instructions for all the R commands you will need for the course, but if you have never used R before, expect a learning curve to get up to speed. Please ask me or our TA for help if you need R help.

Course Prerequisites: Stat 102, Stat 112, Stat 431, Econ 2310 or comparable courses.
Experience with R is very helpful but not required.

Academic Integrity: Students are expected to be familiar with and comply with Penn's Code of Academic Integrity, which is available in the Pennbook, or online at https://catalog.upenn.edu/pennbook/code-of-academic-integrity/. I generally have a zero-tolerance policy for cheating, and all violations will result in substantial penalties. The written
part of the project will be submitted via the TurnitIn feature on Canvas. If you use ideas from other people you must provide proper citations. If you have any doubts or questions about what constitutes academic misconduct, please do not hesitate to contact me.

**Moderate use of Generative AI permitted:** You may use generative AI programs (e.g., tools like ChatGPT) to help generate ideas and brainstorm and to help you with your R code. However, you should note that the material generated by these programs may be inaccurate, incomplete, or otherwise problematic. (I have given some of previous midterm questions to ChatGPT and it gets many of the answers wrong. Citing ChatGPT does not make an incorrect answer correct.) Also, to receive full credit, I generally require you to show your work. You may not submit any work generated by an AI program as your own. If you include material generated by an AI program, it should be cited like any other reference material (with due consideration for the quality of the reference, which may be poor). Any plagiarism or other form of cheating will be dealt with severely under relevant Penn policies.