

SC397: Statistical Surveys, Censuses, and Society (Fall 2019)

Professor: Jerzy Wieczorek
Office: Davis 213
Telephone: 859-5841
Email: jawieczo@colby.edu

Office hours: Available at our Moodle course site

Class meetings: MWF 1:00 – 1:50 in Keyes 103

Class materials: Available at our Moodle course site

Textbooks: *Survey Methodology*, **2nd edition** (required)
by Groves, Fowler, et al.
Complex Surveys: A Guide to Analysis Using R (optional)
by Lumley

Course description:

“Revolves around the role of sampling and surveys in the context of U.S. society. We will examine the evolution of census- and survey-taking in the U.S. in the context of its economic, social, and political uses, eventually leading to discussions about the accuracy and relevance of survey responses, especially in light of various kinds of sampling and nonsampling errors. We will also explore links to sampling methods useful for studying wildlife, forests, and other non-human populations. Students will be required to design, implement, and analyze a survey using rigorous, well-motivated methods. *Prerequisite: Statistics 212.*”

In SC212 you learned about the importance of using random sampling to collect data... but you also found that it's not always logistically practical to take a simple random sample. What other sampling methods can help us balance data-collection costs versus statistical precision, and how should our data analyses account for these complex samples? How can we prevent or handle nonresponse and missing data? What are the ethical issues around collecting, storing, and managing sensitive data? We will also discuss the societal role of government-run surveys around the world, including the upcoming 2020 decennial census of the United States. Students will design and carry out a semester-long survey project with applications beyond the classroom.

Course objectives: By the end of the semester, students will be able to:

1. Express specific ways that surveys and censuses influence American institutions and daily life
2. Write questionnaires and design sampling procedures in ways that (a) promote internal and external validity and (b) balance statistical efficiency with implementation costs
3. Use the statistical software R to analyze survey data with design-based and model-based methods, accounting appropriately for the sampling design

4. Write a scientific report that conveys the design, implementation, analysis, and interpretation of a survey & its results
5. Collaborate with teammates to design, implement, analyze, and report the results of a survey

Grades: Your course grade will be determined as follows (subject to change):

Citizenship/discussion/attendance	10%
Homework	20%
Midterm #1	15%
Midterm #2	15%
<u>Final Project</u>	<u>40%</u>
Total	100%

Moodle: I will be posting all HW assignments on our Moodle site along with hints, schedule updates, and other important class-related info. Check the Moodle site early and often at <https://moodle.colby.edu>

Class sessions: Regular attendance is essential and in-class participation is expected – please take it seriously. Everyone will receive a grade for in-class participation that will be incorporated into your final grade. You need to attend class and you need to be prepared for class. Check the Moodle site for reading assignments prior to class meetings. At times, class will involve activities based on these readings. This class brings together students with varying backgrounds in science, math, and statistics, so it is especially vital that you raise questions in class when you don't understand a particular topic, or if you have a perspective on a topic that you feel would contribute to class discussion. Please bring your textbook with you to class; we will occasionally work through problems in small groups.

Citizenship: Much of the work done for this course relies on working with others. You are expected to participate fully in both class and in-group assignments. You will be graded on this both by the instructor and by the other members of your group. Occasionally I will ask you to fill out peer citizenship assessment forms for your group. These evaluations are confidential and anonymous in the sense that only I will see the results; other students in the class will not. I will, however, provide feedback if necessary. The purpose of these assessments is only to ensure that everyone is participating fully and working together, NOT to assess someone's strengths and weaknesses with the class material.

Homework: Readings and selected problems will be assigned regularly. Some homework problems will require extensive use of the computer package RStudio. Although you are encouraged to discuss problems with each other, I expect each person to submit his or her own work. You may choose to work with a partner on data analyses and thus produce the same computer output – that's okay, but I want written interpretations and responses to be

in your own words. And remember, an answer to a statistics question is almost never just a number – provide thoughtful explanations and conclusions!

Homework due dates and times will be posted on Moodle. Turn in hard-copy homework assignments at my office by 4pm on the due date. ***Late homework will reduce the student's grade substantially (by 25% of the maximum score for each day late), and very late homeworks may not be accepted, in order to let me grade everyone's work quickly and give you timely feedback. The professor will not grant extensions except for documented illness or other extreme circumstances.***

Exams: There will be two mid-term exams. These exams will focus on your ability to express an understanding of statistical concepts, to interpret the output of statistical software, and to engage in statistical thinking on open-ended questions. The exams may involve both in-class and take-home components. See our Moodle page for the two exam dates. **You must meet with me within the first two weeks of classes if you need to reschedule an exam date or request accommodations.**

Data analysis project: Your final exam will consist of a complex-survey design and data analysis project that you will work on with 1 to 3 other classmates. This project will be substantially more sophisticated than in SC212, requiring a complex sampling design; rigorous questionnaire testing and data-management planning; post-collection processing and adjustment for nonresponse; etc. Teams will be expected to carry out a study with real applications beyond our classroom. For example, you may find a sponsor/client with a real question they need answered; or choose a broadly-interesting topic and undergo IRB review so that the anonymized dataset you collect can be publicly released. Your grade will be based on project assignments throughout the semester, including a poster session during our final exam timeslot.

Software: We will be using the software package RStudio to analyze data. You can access RStudio via the web at:

<https://rstudio.colby.edu>

I strongly recommend that you use the webserverd version of R to ensure that we are all working off of the same version. However, if you wish to install R and RStudio on your own machine, you can first download R (the programming language itself) here:

<https://cran.r-project.org/>

After installing R, you can download RStudio (a development environment for R) here:

<https://www.rstudio.com/>

In addition, for exams you will also need a scientific calculator (one that can at least do square roots, logs, and exponential functions).

Academic accommodations:

I am available to discuss academic accommodations that any student with a documented disability may require. Please note that you will need to provide a letter from

the Dean of Studies Office documenting your approved accommodations. *Please meet with me within two weeks of the start of the semester to make a request for accommodations so that we can work together with the College to make the appropriate arrangements for you.*

Kate McLaughlin, Associate Director of Access and Disability Services (kmclaugh@colby.edu) is the primary contact for accommodations and any questions related to educational testing and documentation.

Sexual misconduct/Title IX statement:

Colby College prohibits and will not tolerate sexual misconduct or gender-based discrimination of any kind. Colby is legally obligated to investigate sexual misconduct (including, but not limited to sexual assault and sexual harassment).

If you wish to speak confidentially about an incident of sexual misconduct, please contact Colby Counseling Services ([207-859-4490](tel:207-859-4490)) or the Director of the Gender and Sexual Diversity Program, Emily Schusterbauer ([207-859-4093](tel:207-859-4093)).

Students should be aware that faculty members are considered responsible employees; as such, if you disclose an incident of sexual misconduct to a faculty member, they have an obligation to report it to Colby's Title IX Coordinator. "Disclosure" may include communication in-person, via email/phone/text, or through class assignments.

To learn more about sexual misconduct or report an incident, visit <https://www.colby.edu/sexualviolence/>

Academic honesty & consequences for academic dishonesty:

Honesty, integrity, and personal responsibility are cornerstones of a Colby education and provide the foundation for scholarly inquiry, intellectual discourse, and an open and welcoming campus community. These values are articulated in the Colby Affirmation and are central to this course. Students are expected to demonstrate academic honesty in all aspects of this course.

Academic dishonesty includes, but is not limited to: plagiarism (including quoting sources without " " around the borrowed words and a citation); presenting another's work as one's own; buying or attempting to buy papers or projects for a course; fabricating information or citations; knowingly assisting others in acts of academic dishonesty; violating clearly stated rules for taking an exam or completing homework; misrepresentations to faculty within the context of a course; and submitting the same work, including an essay that you wrote, in more than one course without the permission of instructors.

Academic dishonesty is a serious offense against the college. Sanctions for academic dishonesty are assigned by an academic review board and may include failure on the assignment, failure in the course, or suspension or expulsion from the College.

For more on recognizing and avoiding plagiarism, see: <https://libguides.colby.edu/avoidingplagiarism>

For resources and information on academic integrity, see: <https://www.colby.edu/academicintegrity/>

The Colby Affirmation:

Colby College is a community dedicated to learning and committed to the growth and well-being of all its members. As a community devoted to intellectual growth, we value academic integrity. We agree to take ownership of our academic work, to submit only work that is our own, to fully acknowledge the research and ideas of others in our work, and to abide by the instructions and regulations governing academic work established by the faculty. As a community built on respect for ourselves, each other, and our physical environment, we recognize the diversity of people who have gathered here and that genuine inclusivity requires active, honest, and compassionate engagement with one another. We agree to respect each other, to honor community expectations, and to comply with College policies. As a member of this community, I pledge to hold myself and others accountable to these values.

<https://www.colby.edu/catalogue/2014/07/11/colby-affirmation/>

Tentative weekly schedule:

Week	Dates	Topics	Readings	Friday deadlines
1	Sept 4, 6	Introduction	Groves 1	Teammate picks
2	Sept 9, 11, 13	Elements of sampling, Survey quality	Groves 2, 3	Project topics
3	Sept 16, 18, 20	Survey modes, Ethics	Groves 5, 11	Target pop, mode
4	Sept 23, 25, 27	Statistics of sampling	Groves 4	Revisions, IRB
5	Sept 30, Oct 2, 4	Questionnaires	Groves 7, 8	Design, Qs, n
6	Oct 7, 9, 11	Review		Midterm exam 1
7	Oct 14, 16, 18	Stratified sampling	Groves 4	Pretest Qs
8	[fall recess], Oct 23, 25	Cluster sampling, Nonresponse	Groves 4, 6	Final IRB
9	Oct 28, 30, Nov 1	Post-survey processing	Groves 10	[data collection!]
10	Nov 4, 6, 8	Variance estimation	Groves 10	Project report 1
11	Nov 11, 13, 15	Review	Groves 12	Midterm exam 2
12	Nov 18, 20, 22	Special topics		Project report 2
13	Nov 25, [Thanksgiving]	Special topics, break		
14	Dec 2, 4, 6	Special topics, feedback		Draft final report
Exams	TBD [Dec 11-16?]	Feedback		Final presentation

Possible special topics: Training and managing interviewers. Bias in Big Data. History of the US Census. Capture-recapture methods. Differential privacy. Small area estimation. Combining surveys with auxiliary data, pooling poll results, etc. Cross-validation with survey data. Visualizing survey-weighted data. Record linkage and deduplication. Jerzy's applied work with Hague, PPI, SAIPE. [...or other topics as requested.]

Tentative project schedule:**I.0 Request team/teammate preferences (Fri Sept 6)**

Create a team of 2-4 members, and have one person email me your names. Or email me your topic preferences and schedule constraints, and I will assign you to a team.

I.1 Propose topics (Fri Sept 13) [also see instructions on Team Working Agreement]

Identify two survey projects that your team could implement over the course of the semester. (E.g. an on- or off-campus survey of people, a survey of things such as library books, accounting records, cars, buses, etc.).

One topic should be chosen from the list of **community-partner projects** through the Office of Civic Engagement. The other topic should be **your own idea**, although you are welcome to imagine potential partners/clients. See the separate handout on project ideas and proposals, for examples and general expectations.

For each topic, each team should submit a one- to two-page proposal. Each proposal should begin with a title that indicates the topic and the names of every team member. The section headings of the proposal should include the following three items.

A. Why is this topic interesting? Why does this survey need to be done now? Who is the sponsoring partner (or is there another client for whom you might do the survey)?

B. What research question(s) do you propose to study? Give a brief answer that is understandable by a non-statistician. These are not actual survey questions, but well-defined research questions.

C. What research has already been done on the topic or on the theoretical construct of central importance to your topic? What could be learned from survey results? Each group member should locate and review 1 relevant item of research (e.g., article, report, book, etc.). For each item, provide:

- Full author name(s), date, title, where it appeared (book title and publisher, if it appeared in a book; magazine title, date, page numbers, if it appeared in a magazine; website title and web address if it appeared online, etc.)

- A sentence or two summarizing the item and what its connection is to the survey you are proposing.

- The name of the team member who found this item.

There is no restriction on what kind of surveys you propose, except that they should both be interesting and doable by your team, in this semester.

I.2 Define the target population, sampling frame, mode of data collection, & variables (Fri Sept 20)

For each of the two topics you proposed, each team should add the following sections to their proposal, making it 3-5 pages in all:

D. What is the target population? For what population(s) do you wish to make inferences? What are the observation units?

E. What is the sampling frame? What are the sampling units?

- How does the target population differ from the sampling frame, for your survey?

- What possible non-sampling errors could arise in the survey you plan to conduct?

- Explain each possible error, how it could occur, and how you suggest tackling it.

F. What is the mode of data collection? How do you plan to carry out the survey (e.g., by telephone, e-mail, in-person, etc.) and why?

G. What variables do you propose to measure?

I.3 Write final proposal; draft IRB application (Fri Sept 27)

H. On the basis of feedback to your submission for Parts I.1 and I.2, choose a final survey topic, and update or revise your answers to (A) through (G) above. Submit the revised proposal.

If your team's project involves human subjects/respondents, attach the following to the final, revised proposal:

I. Fill out the IRB form (available on the course webpage). Submit to Jerzy, not to Colby's IRB.

J. How do you plan to protect the privacy and assure the confidentiality of respondents? Talk about data collection, protection, and disclosure.

II.4 Sampling scheme and question design (Fri Oct 4)

K. Decide on a sampling scheme (e.g., SRS, Stratified random sample, etc.) and explain why you chose it.

L. Write a questionnaire with 10-30 questions. Up to approximately 1/3 of these can be background or demographic questions and the rest should be directly related to the research questions you will try to answer with your survey. NOTE: If your survey involves observations instead of asking people questions, then instead you should carefully describe your observation protocol. That is, list 10-30 things you will always look for when observing each unit in your survey.

M. Give some idea of the sample size you will require and how you arrived at this number (talk about the margin of error for inferences you want to make).

II.5 Pretest and revise questionnaire (Fri Oct 18)

N. Pretest a revised version of your questionnaire (or observation protocol) on a group of possible respondents/units.

O. Report on the specification of and results from the pretest, and any redesign of the questionnaire (or observation protocol) that may be required.

II.6 Final IRB and project plan (Fri Oct 25)

P. Submit a final revised version of (I) through (O) for the final IRB form. Again, submit to Jerzy, not Colby's IRB.

Q. Develop and report on your schedule for implementing the survey, including plans for nonresponse followup, data analysis, and preparation of a final report.

Once this final version of the survey plan is approved, you will be able to begin implementation.

III.7 In-class progress report 1 (Fri Nov 8)

III.8 In-class progress report 2 (Fri Nov 22)

Prepare to present short (15-20 minute) presentations on data collection and preliminary analysis.

III.9 Rough draft of final report/deliverables (Fri Dec 6)

III.10 Final presentations (exam week, date TBD)

Final report presentations will take place during exam week. If your team had an outside partner, you will also be responsible for delivering the final report, dataset, and any other deliverables to the client during exam week.