STAT/BIOST/CSSS 529: Sample Survey Techniques, Spring 2017

Syllabus

Instructor: Sahar Z Zangeneh, PhD
Office: M2- C118 Arnold Building FHCRC
Phone: 667-7049 (Fred Hutchinson)
Email: saharzz@fhcrc.org

Time and Place: Tuesdays, Thursdays 10:30 - 11:50 am LOE 101
Office hours: Tuesdays 9:15-10:15 am, F-646 HSB
Class Web Pages: https://canvas.uw.edu/courses/

Teaching Assistant: Jun Hwang
Office: H-669 Magnuson Health Sciences Center
Email: junhwang@uw.edu
Labs: Wednesdays 2-3:30 pm, I-132 ("i" wing) HSB

Course overview:

The course will cover applied analysis of complex survey data, with applications in the health and social sciences and aims to provide guidance on correct application of modern techniques for design-based analysis of complex sample survey data.

The emphasis of this course is primarily on breadth, and in that spirit we aim at minimizing the use of mathematics without sacrificing accuracy. The course will broadly cover survey sampling, analysis of data from complex surveys including descriptive and analytical inference, and the design and analysis of two-phase samples and sampling for rare events. Time permitting we will discuss model-based analysis of survey data and other special topics.

TA lab sections

The TA will hold regular weekly labs/office hours with the aim to reinforce the material covered in lectures. The labs will include hands-on experience with R, going over some mathematical derivations of important formulae, and providing a forum for discussing homework solutions, assigned articles and additional examples of sample survey design and analysis. He will also maintain a discussion board on canvas

Prerequisites

Students must have taken a graduate-level introductory course in applied statistics, and a regression modeling course is recommended. Knowledge of R would be helpful.

Assessment:
Homework (~6 total): 40%
Midterm exam: 30% (tentatively scheduled for May 2nd)
Quiz: 5% (tentatively scheduled for May 25th)
Final project: 25% (proposals due on April 27th and presentations scheduled for May 30 and June 1 2017)

Grading: Graded (3 credits)

Learning Objectives

After successfully completing this course, students should ordinarily expect to be able to:

- Determine when a survey has been well designed, and recognize what could go wrong if when it is poorly designed
- Define a probability sample, explain its importance in statistics, and determine when probability sampling is used
- Define common features of complex surveys, e.g., strata, clusters, and unequal sampling probabilities and explain how they affect the cost of the survey and precision of estimates
- Compute summary statistics and fit regression models to data from complex surveys using R
- Explain why assumptions about the distribution of data are not relevant to standard design-based survey inference and what criteria are relevant for choosing summary statistics and models
- Define post-stratification and raking, and explain how they can increase precision
- Describe some strategies for mitigating the bias from non-response
- Explain the advantages and disadvantages of including sampling weights in a regression model
- Describe case-cohort, case-control, and two-phase case-control designs for sampling from a cohort, and how data from these designs can be analyzed

Computing Software: R

R is a freely available computing package. It is available on departmental computers. It can also be downloaded on your personal computer from http://cran.r-project.org/

Text: There are several good textbooks in Survey Sampling. We will be using Lumley (2010) as the main resource, which students would need to purchase. Heeringa, West and Berglund (2010) and Lohr (2010) are also two excellent books that we will be using as secondary resources; students are however not required to purchase these books. All books are requested to be put on short loan at the Math library (3rd floor of Padelford).

Important Notes:
1. Class material, including lecture notes, homework assignments, and other course-related information will be posted on the webpage. Printed course material will <u>not</u> be provided by the instructor. Please check the webpage regularly and print/copy the notes
2. The instructor reserves the right to modify the course plan and the syllabus, as needed
3. Questions and discussions are welcome, and encouraged throughout the class; keep in mind that if there is something that is not clear to you, it most likely is unclear to others as well

Homework

Weekly homework will be assigned at the beginning of the quarter, and gradually taper off towards the end of the quarter to allow enough time for you to work on your projects. They will typically be posted on Thursdays and collected online the following Friday.

- I encourage you to work on the homework assignments with each other in small groups, however, each student is required to prepare and submit their own solution and write-up
- Aim to resolve all technical questions or problems you might have with running software a few days before an assignment is due
- You need to submit your homework electronically on Canvas, which will remain open until midnight on Fridays

- Homework assignments that are not handed in on time will receive zero points (except in cases of documented emergency)
- Please type up your homework assignments using a text editor and submit as a pdf file
- Include the R code you used for an assignment in the appendix at the end of your write up

Important Notes:

1. Class material, including lecture notes, homework assignments, and other course-related information will be posted on the webpage. Printed course materials will not be provided by the instructor. Please check the webpage regularly and print/copy the notes.
2. Questions by e-mail are welcome. They will often be answered quite quickly, but this is not guaranteed. In particularly, I don't always check e-mail over weekends
3. The instructor reserves the right to modify the course plan and the syllabus as conditions require.
4. Questions and discussions are welcome, and encouraged throughout the class; keep in mind that if there is something that is not clear to you, it most likely is unclear to others as well.

Tentative Course Outline:

1. Introduction and examples

2. Simple random sampling, stratified random sampling, cluster sampling

3. Ratios and quantiles

4. Unequal probability sampling

6. Design

7. Linear regression

8. Logistic (and other) regression

9. Resampling and replicates

10. Post-stratification and calibration

12. Two-phase designs

13. Missing data

14. Special topics (time permitting)

Students with Disabilities

If you would like to request academic accommodations due to a disability, please contact Disabled Student Services, 448 Schmitz, 543-8924 (V/TTY). If you have a letter from Disabled Student Services indicating you have a disability that requires academic accommodations, please present the letter to me so we can discuss the accommodations you might need for this class.

Reference List (in order of relevance)

1. Lumley, Thomas. Complex surveys: A guide to analysis using R. Vol. 565. John Wiley & Sons, 2010 (required textbook)
2. Heeringa, Steven G., Brady T. West, and Patricia A. Berglund. Applied survey data analysis. CRC Press, 2010
3. Lohr, Sharon. Sampling: design and analysis, 2nd edition. Cengage Learning, 2009
4. Cochran, William G. Sampling techniques. John Wiley & Sons, 2010
5. Valliant, Richard, Jill A. Dever, and Frauke Kreuter. *Practical tools for designing and weighting survey samples*. New York, NY: Springer, 2013
6. Skinner, Chris J., David Holt, and TM Fred Smith. Analysis of complex surveys. John Wiley & Sons, 1989
7. Särndal, Carl-Erik, Bengt Swensson, and Jan Wretman. Model assisted survey sampling. Springer, 2003
8. Korn, Edward L., and Barry I. Graubard. Analysis of health surveys. Vol. 323. John Wiley & Sons, 2011
9. Valliant, Richard, Alan H. Dorfman, and Richard M. Royall. Finite population sampling and inference: a prediction approach. John Wiley & Sons, 2000
10. Chambers, Raymond L., et al. Maximum likelihood estimation for sample surveys. Vol. 125. CRC Press, 2012
11. Little, Roderick JA, and Donald B. Rubin. Statistical Analysis with Missing Data, 2nd edition. John Wiley & Sons, (2002
12. Wolter, Kirk. Introduction to variance estimation. Springer, 2007