SURVMETH 988

Probability and Non-probability Sampling Methods

Summer 2018

CLASS MEETINGS:

Lecture: 9:00-12:00 PM Monday-Friday, July 16-20, 2018
TBD, University of Michigan

Lab: 1:00-4:00 PM Monday-Friday, July 16-20, 2018
TBD, University of Michigan

INSTRUCTORS

James M. Lepkowski
Room 4010 ISR, 426 Thompson Street, University of Michigan
Telephone: 734 678 6964 (cell)
Email: jimlep@umich.edu

Sunghee Lee
Room 4086 ISR, 426 Thompson Street, University of Michigan
Telephone: 734 615 5264 (office)
Email: sungheel@umich.edu

ASSISTANT

Jodi Holbrook
Room 4050 ISR, 426 Thompson Street, University of Michigan
Telephone: 734 764 6595
Email: hjodi@umich.edu

COURSE CONTENT

Probability sampling methods have been and continue to be the standard against which sampling designs are compared. Non-probability sampling methods have been widely used in the past, but greater interest today has arisen because they are cheaper and faster. Additionally, there are population subgroups for whom probability sampling quickly loses its feasibility. However, non-probability sampling methods have unknown properties and potentially larger bias than probability methods.

This course, unlike other sampling methods courses, compares these two types of sampling methods. The premise is that non-probability sampling cannot be discussed meaningfully without understanding probability sampling first.

The course thus first examines probability sampling techniques and their properties, including simple random selection, systematic selection, cluster sampling, stratified sampling, and probability proportionate to size selection. Issues of weighting to compensate for unequal chances of selection and variance estimation for calculating confidence intervals are also

examined.

Then non-probability sampling techniques are examined, from online recruited panel-based and river sampling to quota sampling and respondent-driven sampling, and techniques. The currently known properties of these sampling techniques are discussed, as are the assumptions needed to calculate estimates of population characteristics and the reliability of those estimates.

In addition to the lectures on theoretical/conceptual parts of sampling as discussed above, we will offer practical applications of different approaches through lab sessions. These sessions to be held after each class will combine R programming and group discussions on the topics that need to be considered when implementing various sampling approaches. Hands-on examples of frame preparation, sample draws, post-survey adjustments and analysis specific to design will be provided and discussed.

The course aims to provide with sufficient background to understand the reasons for using probabilities in sample selection, the consequences of not using probability sampling, and the nature of non-probability sampling techniques in order to determine strengths and weaknesses of commonly used techniques.

COURSE CANVAS WEB SITE

All participants have access to the course web site on the University of Michigan Canvas system. The site contains the course syllabus, lecture notes, readings, video recorded lectures, in-class exercises, homework assignments, and other materials. Participants will be given login information prior to the start of the course. They are expected to have logged in successfully before the first class session, reviewed the syllabus, and watched six video recorded lectures (see the Syllabus below) prior to the first class session. Staff assistance is provided if participants are having difficulty logging in.

COURSE LOGISTICS

This is a one-week course with in-person participation. The textbook is short and relatively easy to read. It is supplemented in the first three days with video recordings of lectures on key topics in probability sampling. Video recordings are available through the course web site. The video recordings for each of the first three days are approximately two and one-half hours in length, and are to be viewed prior each of the first three class sessions.

Class sessions consist of discussion of readings, video lectures (the first three days), in-class group exercises, and homework assignment solutions. Readings other than the textbook, group exercises, solutions to group exercises, and homework assignments are be distributed through the course web site. Homework exercises are collected via the course web site as well.

The class sessions are recorded daily using the Bluejeans system and uploaded to the course web site Media Gallery. The Bluejeans system can also be used for remote access to class sessions. Thus, it is possible to take this course entirely online. Please contact the instructors if interested in the online option.

ATTENDANCE

Participants anticipating missing a class session due to illness or other reasons should

notify the instructors via email at least 30 minutes before the session that they will miss. Each student is permitted one excused absence. The course average score is reduced by 10% for each unexcused absence.

HOMEWORK AND GRADING

There are three 100-point homework assignments. The first two assignments address probability sampling techniques. The third 100-point assignment (the "project") is a critique of assigned readings on non-probability sampling techniques. The final grade will be a weighted average score adjusted for unexcused absences.

Homework assignments are downloaded from the course web site Assignment button. Completed assignments may be handwritten or typed, but all must be uploaded to the course web site Assignment button in a single .pdf format file. Name and homework number should be given at the top of the first page, and page numbers included at the bottom of each page. Course assignment files should be named following the convention 'Surname First-Initial HW #.pdf' (for example, 'Lepkowski J HW 1.pdf'). The marked completed assignment, along with a copy of solutions for probability sampling assignments, are returned via the course web site.

Study groups are useful and encouraged for preparing answers to homework exercises. Group answers are not acceptable. That is, each participant must submit an individual homework solution.

ACADEMIC INTEGRITY

We expect participants to follows University of Michigan ethical standards presented in the academic integrity policies in the University of Michigan Standard Practice Guide (http://spg.umich.edu/pdf/303.03.pdf). Violation of the policy is a serious matter and may lead to a failing mark on a submission or for the entire course.

STUDENT COURSE EVALUATION

Students will at the end of the course be sent a link to an online student course evaluation. All students are strongly urged to complete the online course evaluation.

VIDEO LECTURES ON PROBABILITY SAMPLING TECHNIQUES

There are 14 recorded lectures available on the course site Media Gallery. The recorded lectures total seven hours and 28 minutes. The assigned viewing each of the first three days (see the Syllabus below) is approximately two and one-half hours each:

01) Where does sampling fit in? (11 minutes 49 seconds)
02) Course topics (22 minutes 23 seconds)
03) Simple random sampling (21 minutes)
04) Historical perspectives (24 minutes 58 seconds)
05) Element sampling (44 minutes 44 seconds)
06) Systematic sampling (20 minutes 45 seconds)

07) Cluster sampling, part 1 (33 minutes)
07) Cluster sampling, part 2 (30 minutes 27 seconds)
08) Two-stage cluster sampling (25 minutes 23 seconds)
09) Probability proportionate to size sampling (36 minutes 54 seconds)
10) Stratified random sampling (38 minutes 44 seconds)
11) Frames (42 minutes 36 seconds)
12) Weighting, part 1 (36 minutes 18 seconds)
12) Weighting, part 2 (27 minutes 34 seconds)
13) Variance estimation (26 minutes)
14) Conclusion, probability sampling techniques (5 minutes 25 seconds)

TEXTBOOK

The principal text for the course will be *Introduction to Survey Sampling* by Graham Kalton (Sage Publications, Beverly Hills, 1983).

ASSIGNED READINGS ON NON-PROBABILITY TECHNIQUES

The following readings are available on the course Canvas site Files folder "Non-probability techniques reading":

Pew Research Center (2016). *Evaluating Online Nonprobability Surveys. Vendor choice matters; widespread errors found for estimates based on blacks and Hispanics.* Available from http://www.pewresearch.org/files/2016/04/Nonprobability-report-May-2016-FINAL.pdf

Heckathorn, D.D. (1997). "Respondent-driven sampling: a new approach to the study of hidden populations." *Social Problems* 44: 174–199.

Lee, S., Suzer-Gurtekin, Z.T., Wagner, J. and Valliant, R. (2017). "Total Survey Error and Respondent Driven Sampling: Focus on Nonresponse and Measurement Errors in the Recruitment Process and the Network Size Reports and Implications for Inferences." *Journal of Official Statistics*, 33(2), 335-366. DOI: https://doi.org/10.1515/jos-2017-0017

Lee, S., Wagner, J., Valliant, R., and Heeringa, S. (2014). "Recent Developments of Sampling Hard-to-Reach Populations: An Assessment." In Tourangeau, R., Edwards, B., Johnson, T., and Wolter, K. (eds.) pp. 424-444, *Hard to Survey Populations.* Cambridge, UK: Cambridge University Press

Semaan, S., Lauby, J., & Liebman, J. (2002). "Street and Network Sampling in Evaluation Studies of HIV Risk-Reduction Interventions." *AIDS Review, 4,* 213-223.

Wang, W., Rothschild, D., Goel, S., & Gelman, A. (2015). "Forecasting Elections with Non-representative Polls." *International Journal of Forecasting*, 31(3), 980-991.

READING FOR PROJECT PAPER

The following readings are available on the course Canvas site Files folder "Project paper reading":

Raymond, H. F., Rebchook, G., Curotto, A., Vaudrey, J., Amsden, M., Levine, D., & McFarland W. (2010). "Comparing Internet-Based and Venue-Based Methods to Sampling MSM in

the San Francisco Bay Area." *AIDS and Behavior*, *14*, 218-224.

Kendall, C., Kerr, L., Gondim, R., Werneck, G., Macena, R., Pontes, M., Johnston, L., Sabin, K., & McFarland, W. (2008). "An Empirical Comparison of Respondent-Driven Sampling, Time Location Sampling, and Snowball Sampling for Behavioral Surveillance in Men Who Have Sex with Men, Fortaleza, Brazil." *AIDS and Behavior*, *12(Suppl 4)*, 97-104.

SYLLABUS

Monday, July 16

Learning Objectives: Course introduction and overview, historical accounts on sampling. Simple random & cluster sampling.
Reading: Kalton, Chapters 1, 2, 3, and 4
Assignment: HW 1 assigned
Videos: (1), (2), (3), (4), (5), & (6)
Lab: Overview on R; Frame preparation; Drawing simple random sample; Sampling error (bias and standard error); Drawing cluster sampling

Tuesday, July 17

Learning Objectives: One- & two-stage cluster & probability proportionate to size sampling, and stratified sampling.
Reading: Kalton, Chapters 5, 6, 7, and 8
Assignment: HW 1 due; HW2 assigned
Videos: (7 part 1), (7 part 2), (8), (9), & (10)
Lab: Drawing one vs. two stage cluster samples; Probability proportionate to size samples and stratified samples; Estimation by sample design

Wednesday, July 18

Learning Objectives: Frames, frame problems, and remedies through the use of weights. Weighting to compensate for unequal probabilities of selection. Variance estimation. Sampling for rare populations. Venue based sampling.
Reading: Kalton, Chapters 9, 10, 11, and 12; Lee et al., 2014
Assignment: HW 2 due; Project reading assigned
Videos: (11), (12 part 1), (12 part 2), (13), & (14)
Lab: Inspecting frames vs. sample; How to compute selection weights; Nonresponse adjustment; Post-stratification; Putting all things together: inference using complex sample surveys

Thursday, July 19

Learning Objectives: Time & location sampling. Snowball & network sampling, and respondent-driven selection.
Reading: Kalton Chapters 13 and 14; Heckathorn, 1997; Lee et al., 2017; Semaan et al., 2002
Homework: Read project papers
Lab: Data structure for respondent driven sampling; Inferences using respondent driven sampling

Friday, July 20

Learning Objectives: Web-based panel and river sampling. Quota sampling. Convenience and judgment sampling.
Reading: Pew Research Center, 2016, and Wang et al., 2015.
Homework: Project paper due
Lab: Propensity score adjustment for nonprobability samples; Comparison on estimates from probability vs. nonprobability samples; Web survey panels