# Methods and Theory of Sample Design

*STATISTICS 580/BIOSTATISTICS 617/SURVMETH 617/SOCIOLOGY 717*

# Sampling Theory

*SURV 440*

# Fall, 2013

## INSTRUCTOR

Jim Lepkowski
jimlep@umich.edu
734-678-6964

## CLASS MEETINGS

1:10-2:30 PM Monday & Wednesday
Ann Arbor, MI: G300 ISR-Perry; College Park, MD: 1208 Lefrak Hall

## OFFICE HOURS

Chat room on CTools or by appointment (in person, by telephone, or by Skype)

## COURSE CONTENT

Methods and Theory of Sample Design is concerned with the theory underlying the widely used practical methods of survey sampling. It covers the basic techniques of simple random sampling, stratification, systematic sampling, cluster and multi-stage sampling, and probability proportional to size sampling; methods of variance estimation for complex sample designs, including the Taylor series expansion method, balanced repeated replications, and jackknife methods; and several specialized topics, including stratification and subclasses, multi-phase or double sampling, ratio and regression estimation, selection with unequal probabilities without replacement, non-response adjustments, and imputation. The course examines the nature of the techniques, practical applications, and the theory supporting the methods.

## COURSE ORGANIZATION, HOMEWORK, EXAMINATIONS, AND GRADING

The course meets twice per week, starting at 10 minutes past the hour (1:10 PM) and ending at 2:30 PM. Class time is devoted to lecture and discussion of homework and examination problems. Questions during lecture are welcomed.

Course materials are on the course web site in CTools, including lecture notes, homework problems, and online readings. The first lecture notes section will be distributed in class; students are responsible for printing copies of other sections.

Homework assignments are assigned as indicated in the syllabus below. Assignments are due on the date indicated at the start of the class session. The instructor is available in a CTools chat room session the Monday evening before each homework assignment is due, usually 9:00-10:00 PM. Any changes in the chat room schedule will be announced in class and via the course CTools site the week before the assignment is due.

Homework is to be submitted electronically via the course CTools web site 'Assignment tool' as an attachment. Solutions may be handwritten or typed, but submitted in a single .pdf format file, with name and homework number at the top of the first page, and page numbers at the bottom of each page. Files must be submitted with a standard name convention: 'Surname First-Initial HW #.pdf' (for example, 'Lepkowski J HW 1.pdf'). The submitted homework will be marked and returned along with a copy of a solution as an Assignment tool attachment.

Study groups are useful, and encouraged, for preparing answers to homework exercises and the project. Group answers are not acceptable, except for the project; each student must submit individual homework solutions.

Homework assignments are graded on a 100 point scale. Without prior arrangement, homework submitted after the due date will be considered late, and 20 points deducted.

There are two open-book open-notes in-class 100 point examinations, a mid-term and a final. Each cover all material discussed previously in the course. Students are advised to bring a calculator or a laptop to the examinations. Although laptops may be used to aid calculation or review notes, no internet connections of any kind are allowed during the exams.

The final grade for the course is an average of homework and examination scores: final examination contributes approximately 40%, the midterm examination 30%, the homework 30%.

Regular attendance and participation in this class is the best way to grasp the principles and applications discussed. In the event that a class must be missed due to an illness or other compelling reason, a reasonable effort should be made to notify the instructor in advance of the class. If a student is absent on days when exams are scheduled or homework due, the student must make arrangements with the instructor in advance of the class session to reschedule the exam or submit the homework later than the due date.

Video recordings will be made of each class sessions and may be viewed later using 'browser software' later. The link and password to the server where the recordings are stored will be distributed via the course CTools site Announcements tool.

## ACADEMIC CONDUCT

The faculties of the University of Maryland and University of Michigan expect registered students or students taking courses to conduct themselves professionally. Courtesy, honesty, and respect should be shown by students toward faculty, guest lecturers, support staff, and fellow students. Similarly, students should expect faculty to treat them fairly, showing respect for their ideas and opinions and striving to help them achieve maximum benefits from their experience in the school.

Student academic misconduct includes plagiarism, cheating, fabrication, falsification of records or official documents, intentional misuse of equipment or materials, and aiding and abetting the perpetration of such acts. The preparation of reports, papers, and examinations assigned on an individual basis must represent each student's own effort. Reference sources should be indicated clearly. The use of assistance from other students or aids of any kind during a written examination, except when the use of books or notes has been approved by an instructor, is a violation of standards of academic conduct.

## COURSE EVALUATION

Both the University of Maryland and University of Michigan have online course evaluations. All students are strongly urged to use the respective university online course evaluation system at the end of the term. Course evaluations will only be available to other students when at least 70% of registered students complete the evaluation. Students will be reminded to log on toward the end of the semester to complete a course evaluation.

## TEXTBOOKS

There is one main text for the course: S*ampling Techniques, 3rd ed.,* by William G. Cochran (J. Wiley and Sons, Inc., New York, 1977). There are five required online readings available through the CTools Resources Tool:

(1) Blumenthal, Mark. "'Unskewed Polls' Critics Miss Basics of Party Affiliation," in *Huffmaster Post*, October 1, 2012.

(2) Kish, L. "Selection With Probabilities Proportional to Size Measures," Chapter 7 in *Survey Sampling*, pp. 217-247. New York: Wiley and Sons, 1965.

(3) Rust, K, and Rao, J.N.K. "Variance Estimation for Complex Surveys Using Replication Techniques," *Statistical Methods in Medical Research*, Vol. 5, 1996, pp. 283-310.

(4) Kish, L. "Biases and Nonsampling Errors," Chapter 13 in *Survey Sampling*, pp. 509-571. New York: Wiley and Sons, 1965.

(5) Sarndal, C-E, "The Calibration Approach in Theory and Practice," *Survey Methodology*, Vol. 33, No. 2, 99-119.

Optional reading is also available in *Survey Sampling* by Leslie Kish (J.W. Wiley & Sons, New York, 1965). The Course Syllabus below shows the specific sections and dates when the required and optional readings are to be completed.

**COMPETENCY**

The student will learn and will be tested on the following competencies:

1. Learn the meaning and application of conditional expectation and conditional variance in deriving expressions for sampling variance of means and proportions under a variety of finite population sampling techniques.
2. Understand the properties of various sampling techniques including simple random sampling, stratification and stratified random sampling, cluster sampling, systematic sampling, two-stage sampling, multistage sampling, probability proportionate to size sampling, stratified multistage sampling, and two-phase sampling.
3. Understand how sampling variance for means and proportions are derived for each of the sampling techniques in the course, and derive procedures for complex statistics using the Taylor series expansion.
4. Learn how sampling variance is estimated for complex sample surveys, including using such techniques as balanced repeated replication, jackknife repeated replication, and the Taylor series expansion approximation, and when to apply each in practice.
5. Understand how calibrated and model assisted estimation can be used to improve the precision of survey estimates.
6. Learn how nonresponse can affect survey estimates, and what techniques can be used to reduce nonresponse and compensate through weights and imputation.

**COURSE SYLLABUS**

| Date | Topic | Readings[a] | Homework[b] |
|---|---|---|---|
| *Sept 4* | Course introduction. Statistical tools. | C 1; K 2.0-2.2 | -- |
| *9* | Simple random sampling. | C 2; K 2.3-2.5 | -- Chat |
| *11* | Simple random sampling. | C 3; K 2.6-2.9 | Homework 1 |
| *16* | Simple random sampling. Stratification. | C 4; K 3.1-3.3 | -- |
| *18* | Stratified sampling. | C5 | -- |
| *23* | Stratified Sampling. | C 5A; K 3.4-3.5 | -- Chat |
| *25* | Stratified Sampling.[c] | -- | Homework 2 |
| *30* | Stratified Sampling | -- | -- |
| *Oct 2* | Stratified Sampling. | O(1) | -- |
| *7* | Stratified Sampling | -- | -- Chat |
| *9* | Stratified Sampling. Cluster sampling. | C 9; K 5.1-5.2 | Homework 3 |
| *14* | ***No class: Michigan Fall Break.*** | -- | -- |
| *16* | Cluster sampling. | -- | -- |
| *21* | Systematic sampling | C 9A; K 6.1-6.3 | -- Chat |
| *23* | ***Midterm exam; 1:10-2:30 PM.*** [c] | -- | -- |
| *28* | Systematic sampling. Two stage sampling. | C8; K 4.1-4.2 | -- |
| *30* | Two stage sampling. | C 10; K 5.3-5.4 | -- |
| *Nov 4* | Unequal sized cluster sampling I. | C9A; K6.1-6.3 | -- |
| *6* | Ratio & regression estimation.[c] | C6; K 6.6,6.12 | -- |
| *11* | Unequal sized cluster sampling II. | C 11; K 6.4 | -- |
| *13* | Probability proportionate to size sampling. | O(2); K7.1-7.5 | -- |

| | | | |
|---|---|---|---|
| *18* | Probability proportionate to size sampling. | -- | -- [Chat] |
| *20* | Sampling errors for non-linear statistics[c] | C 11; K 4.3-4.4 | Homework 4 |
| *25* | Sampling errors for non-linear statistics | O(3) | -- |
| *27* | Unequal probability w/o replacement | C 11 | -- |
| *Dec 2* | Double sampling. | C 12; K 12.1 | -- [Chat] |
| *4* | Unit non-response[c] | C 13 | Homework 5 |
| *9* | Weighting and other compensation. | O(4);K 13.1-13.3 | -- |
| *11* | Calibrated estimators[c] | O(5) | -- |
| *13* | ***Final exam; 4:00-6:00 PM.*** | -- | -- |

[a] Refers to sections of Cochran (C) or Online (O) which should be read before the class session. Sections from Kish (K) are optional.

[b] Assignment due at the beginning of the indicated class period. Problem sets posted on CTools Assignment Tool at least one week before the assignment is due.

[c] Instructor at JPSM, 1208 Lefrak Hall, University of Maryland

[Chat] Chat room sessions are Monday evenings before homework is due, 9:00-10:00 PM, unless specified otherwise. For example, the chat room session for the first homework is Monday, September 9, 9-10 PM.