

SURVEY 625

Applied Sampling
Summer 2010

SURVMETH 612

Methods of Survey Sampling
Spring/Summer 2010

CLASS MEETINGS:

3:00-5:30 PM, Tuesday & Thursday, June 8 – July 29, 2010
Room T-10 Census Bureau, Room 1208 Lefrak and Room G300 ISR-Perry

INSTRUCTOR

James M. Lepkowski
Room 4053 ISR-Thompson (U of Michigan) and Room 1218S Lefrak (U of Maryland)
Telephone: 734/678-6964 (cell) and 301/314-7911 (U of Maryland)
Email: jimlep@umich.edu or jlepkowski@survey.umd.edu

James Wagner
Room G374 ISR-Perry (U of Michigan) and Room 1218S Lefrak (U of Maryland)
Telephone: 734/647-5600 and 301/314-7911 (U of Maryland)
Email: jameswag@umich.edu

ASSISTANTS

Mahmoud ElKasabi
Room 4062 ISR, 426 Thompson Street, University of Michigan
Telephone: 734/647-0421
Email: mkasabi@umich.edu

Hanzhi Zhou
Room 4062 ISR, 426 Thompson Street, University of Michigan
Telephone: 734/647-0421
Email: zhouhanz@umich.edu

Adam Schlecte
Room 4050 ISR, 426 Thompson Street, University of Michigan
Telephone: 734-647-3592
Email: schadam@umich.edu

OFFICE HOURS

By appointment. See also the 'Chat Room' schedule

COURSE CONTENT

Methods of Survey Sampling/Applied Sampling is an applied statistics methods course, but differs from most statistics courses because it is concerned almost exclusively with the *design* of data collection. Little of the analysis of collected data will be discussed in the course. The course will concentrate on problems of applying sampling methods to human populations, since sampling human populations poses a number particular problems not found in sampling of other types of units. The principles of sample selection, though, can be applied to many other types of populations.

The course is presented at a moderately advanced statistical level. While we will not develop mathematical aspects of sampling theory, statistical notation and outlines of some algebraic proofs will be

given. A sound background in applied statistics is necessary, since some algebraic derivations will be presented (although little emphasis will be placed on the derivations). A thorough understanding of the notation and results will be required.

The aims of the course are to teach basic ideas of sampling from an applied perspective and to provide experience with life-like problems. The course will cover the main techniques used in sampling practice: simple random sampling, stratification, systematic selection, cluster sampling, multistage sampling, and probability proportional to size sampling. These methods will be examined further in the context of two particular types of sample designs, area sampling and telephone sampling. The course will also cover sampling frames, cost models, sampling error estimation techniques, non-sampling errors, and compensating for missing data.

The course will be simultaneously transmitted through compressed video technology between the Joint Program in Survey Methodology at the University of Maryland, the Institute for Social Research at the University of Michigan, and the US Census Bureau in Suitland, Maryland. The three-way interactive audio-video system allows those at the transmitting site to see and hear those at the remote site, and those at the remote site to see and hear those at the transmitting site. Three of the class sessions will be taught in College Park; the remaining sessions will be taught from the University of Michigan and transmitted to the remote sites. The expected dates when the instructor is scheduled to be in Maryland or Michigan are indicated in the attached syllabus (see footnote).

Class sessions are twice per week. Class will start at 3:00 PM each day. There will be one 10 minute break, and class will end at 5:30 PM.

Lecture notes and other materials will be presented on projection systems in each room. Students have access to a copy of all materials presented on the projectors through the course web site, although handwritten notes will be inserted in lecture frequently.

All registered students have access to the site through registration (at the University of Michigan) or guest log in (for University of Maryland and Census Bureau students). The web site contains lecture notes, homework problems and related materials, homework solutions, readings, the course project and related materials, discussion items, an email log, and the chat room. Materials posted on the web site will not be distributed in class (except the first class session when paper copies of the syllabus, lecture notes, and first homework problem will be distributed).

HOMEWORK

The homework assignments are of two types, problem sets and a project. Both types are to be turned in by the beginning of the class session when due. The regular problems will be graded on a five level system: check-plus (100), check (90), check-minus (80), late (60), not submitted (0). The 'late' score will be assigned for any assignment turned in after the assigned time and day, without prior permission of the instructors. Homework will be submitted electronically via the course web site as an attachment to the Assignment tool. Students must submit solutions, handwritten or typed, in a single .pdf format file, with name and homework set number at the top of the first page, and page numbers at the bottom of each page. Files must be submitted in a standard name convention: 'Surname First Initial HW #.pdf'. For example, 'Lepkowski J HW 1.pdf'. The submitted homework will be marked electronically and returned via the Assignment tool as an attachment, along with a copy of a homework solution. Study groups are useful, and encouraged. Group answers are not acceptable. Each student must submit individual homework exercise solutions.

The project is a two-stage sampling exercise distributed early in the term, and discussed throughout the course during class sessions. The instructors will assign 4-6 students each to project teams during week three (around June 24). Each team will submit one copy of the project in .pdf format with name 'Team # project.pdf' (for example, 'Team A project.pdf') via the Assignment tool. All students in a

team receive the same base score (maximum 80 points). Each student also completes an evaluation for the other students in the team. The remainder of each student's project grade (maximum 20 points) will be based on the evaluations by fellow students in the team. Students who do not turn in evaluations of other team members will receive zero for the individual component, regardless of team ratings.

EXAMINATIONS, FINAL GRADE, AND ACADEMIC INTEGRITY

There will be a one and one-half hour separate in-class cumulative open book, open notes midterm examination on Wednesday, July 7, 3-4:30 PM. The cumulative, open book, open notes final examination will be held Thursday, July 29, 3-5 PM.

Final grades will be a weighted composite of homework (approximately 30%), class project (approximately 30%), and examination scores (approximately 40%). The instructors may alter the relative weights, depending on overall class performance on each component of the final grade.

The course will abide by ethical standards set at the respective campuses (Census Bureau students are University of Maryland students). The University of Maryland, College Park has a nationally recognized Code of Academic Integrity, administered by the Student Honor Council. This Code sets standards for academic integrity for all undergraduate and graduate students. As a student you are responsible for upholding these standards for this course, and being aware of the consequences of cheating, fabrication, facilitation, and plagiarism. For more information on the University of Maryland Code of Academic Integrity or the Student Honor Council, please visit <http://www.studenthonorcouncil.umd.edu/whatis.html>.

TEXTBOOKS AND ASSIGNED READING

The principal text for the course will be *Survey Sampling* by Leslie Kish (John Wiley and Sons, Inc., New York, 1965). It is available at university bookstores or through online sales. Students may find that the following texts serve as useful supplemental reading to several lecture topics: *Introduction to Survey Sampling* by Graham Kalton (Sage Publications, Beverly Hills, 1983), *Sample Survey Methods and Theory*, Volume 1, by Morris Hansen, *et al.* (New York: John Wiley and Sons, Inc., 1953), and *Sampling Techniques*, 3rd edition, by William G. Cochran (New York: John Wiley and Sons, Inc., 1977).

There are also assigned readings of several papers (see list below). These will be available on the course web site.

- [1] Casady, Robert J., and James M. Lepkowski. "Telephone Sampling," Chapter 15 in *Sampling of Populations: Methods and Applications*, Paul Levy and Stanley Lemeshow. New York: Wiley and Sons, Inc., 1999.
- [2] Rust, K. "Variance estimation for complex estimators in sample surveys," *Journal of Official Statistics*, 1(4) (1985): 381-397.
- [3] Kish, L. and Frankel, M. "Inference from complex samples," *Journal of the Royal Statistical Society, Series B*, **36** (1974): 1 - 37.
- [4] Kalton, G. and Kasprzyk, D. "The treatment of missing survey data," *Survey Methodology*, **12** (1986): 1 - 16.

SYLLABUS

Date	Topic	Readings ^a	HW
June 8	Course introduction. Principles in sample selection. Simple random sampling. (JL)	Kish 1.0-1.7	
9	Chat room, 7:30-8:30 PM, CTools Chat Tool.		

Date	Topic	Readings ^a	HW
10	Simple random sampling. Frame problems. Weights and weighted estimators. (JL)	Kish 2.1-2.7	1
15	Cluster sampling. Two stage sampling. Intra-cluster homogeneity. (JL) ^b	Kish 5.1-5.4	
16	Chat room, 7:30-8:30 PM, CTools Chat Tool.		
17	Unequal sized cluster sampling. Stratified sampling. (JL)	Kish 6.1-6.2, 3.1-3.3	2
22	Sample allocation to strata. Stratification topics. (JL)	Kish 3.4-3.6	
23	Chat room, 7:30-8:30 PM, CTools Chat Tool.		
24	Stratified cluster sampling. Systematic sampling. Project introduction. (JL) ^b	Kish 4.1-4.2, 6.3-6.5	3
29	Complex sampling: weighting, stratification, multi-stage selection. Project discussion. (JL)	Kish 5.5, 11.7	
30	Chat room, 7:30-8:30 PM, CTools Chat Tool.		
July 1	Controlling sample size. Probability proportionate to size selection. (JL)	Kish 7.1-7.3	4
6	Probability proportionate to size selection. PPS problems. PPS selection. Review. (JL)	Kish 7.4-7.5	
7 ^c	Midterm examination, 3:00-4:30 PM. Open book, open notes in-class cumulative examination.		
8	Exam review. Area sampling: two and three stage sampling. Project discussion. (JW).	Kish 9.1-9.5	
13	Area sampling: two and three stage sampling. Telephone sampling. Project discussion. (JW) ^b	Kish 9.6-9.7, [1]	
14	Chat room, 3:30-4:30 PM, CTools Chat Tool.		
15	Telephone sampling. Variance estimation: collapsing & combining strata. Project discussion. (JW)	Kish 4.3-4.4, [2]	5
20	Variance estimation: collapsing & combining strata. Balanced and jackknife repeated replication. Project discussion. (JW)	Kish 14.1-14.3, [3]	
21	Chat room, 3:30-4:30 PM, CTools Chat Tool.		
22	Balanced and jackknife repeated replication. Generalized variances. Software. Project discussion (JW).		6
26	Chat room, 3:30-4:30 PM, CTools Chat Tool.		
27	Total survey error. Response error. Non-observation error. Nonresponse adjustment (JW)	Kish 13.1-13.2, [4]	Project
28 ^d	Nonresponse adjustment. Item missing data. Compensating for	Kish 13.3-	

Date	Topic	Readings ^a	HW
	item missing data. Review. (JW)	13.6	
29	Final examination. 3:00-5:00 PM. Open book, open notes in-class cumulative examination.		

^a Readings are from the textbooks by Kish, or from specified papers.

^b Lecture at JPSM.

^c Extra Wednesday session for midterm exam.

^d Added class session, to allow complete coverage of material before final exam. This class session will, like all others, be recorded so that it can be viewed later by those unable to attend on Wednesday.