Applied Sampling
SURV 626
2 credits/4 ECTS
Fall 2017

Instructor(s)
Raphael Nishimura, raphaeln@umd.edu

Short Course Description
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 of 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 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. The course will also cover sampling frames, cost models, and sampling error estimation techniques.

Course and Learning Objectives
By the end of the course, students will…

- understand the basic ideas, concepts and principles of probability sampling from an applied perspective
- be able to identify and appropriately apply sampling techniques to survey design problems
- be able to compute the sample size for a variety of sample designs
- understand and be able to assess the impact of the sample design on survey estimates
- be able to estimate the precision of the survey statistics using different estimation techniques

Prerequisites
The course is presented at an intermediate 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, proficiency in mathematics, including basic algebra, 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 algebraic results will be required.

Class Structure and Course Concept

This is an online course using a flipped classroom design. It covers the same material and content as an on-site course but runs differently. In this course, you are responsible for watching video recorded lectures and reading the required literature for each unit and then "attending" mandatory weekly one-hour online meetings where students have the chance to discuss the materials from a unit with the instructor. Just like in an on-site course, homework will be assigned and graded and there will be a final exam at the end of the course.

Although this is an online course where students have more freedom in when they engage with the course materials, students are expected to spend the same amount of time overall on all activities in the course – including preparatory activities (readings, studying), in-class-activities (watching videos, participating in online meetings), and follow-up activities (working on assignments and exams) – as in an on-site course. As a rule of thumb, for each credit offered by a course, students can expect to spend one hour per week on in-class activities and three hours per week on out-of-class activities over the span of a full 12-week term. This is a 2-credit course that runs for 8 weeks. Hence, the average workload is about 12 hours per week.

Mandatory Weekly Online Meetings

[Day, Time TBD] from [Date TBD] to [Date TBD]

Meetings will be held online through BlueJeans. Follow the link to the meeting sessions on the course website on jpsmonline.umd.edu. If video participation via Internet is not possible, arrangements can be made for students to dial in and join the meetings via telephone.

A student missing a class session due to illness or other reasons must notify the instructor via email at least an hour before the class session. As class participation is part of the grade and the quiz can only be taken as part of the meeting, missing more than a few classes is strongly discouraged.

Grading

Grading will be based on:

- Homework assignments (50% of the grade)
- Quizzes (15% of the grade)
- Participation in discussion during the weekly online meetings, submission of questions via e-mail (deadline: [Day, Time TBD]) demonstrating understanding of the required readings and video lectures and positive contributions on Piazza (10% of grade)
- A final open-book online exam (25% of grade)

Homework assignments

The homework assignments will involve small-scale, sample design problems that will require

you to identify and apply the methods and techniques covered in the lectures and assigned readings. The questions will require mathematical calculations and you will be asked to select samples using different sampling schemes. Although some examples of statistical software will be provided, none of the homework problems will require their use, and the assignments should preferably be solved by hand, with a calculator, or in a spreadsheet, so that you can have a more robust understanding of the concepts being applied in these exercises. Use the homework assignments as an indicator of your progress in this course.

Homework solutions should be submitted electronically via the course web site Assignment tool as an attachment. You 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. Handwritten versions must be fully legible: if the instructor cannot read the homework it will be returned ungraded. Files must be submitted in a standard name convention: 'Surname First Initial HW#.pdf'. For example, 'Nishimura R HW1.pdf'. Homework problems will be graded on a 100-point scale. The submitted homework will be marked electronically and returned via the Assignment tool, along with a copy of the homework solution.

Homework assignments are due the [Day TBD] after the online meeting (see schedule syllabus below). Late homework will not be accepted, except in case of emergencies, which should be reported to the instructor in advance through a request made in writing by email no less than 24 hours before the homework is due, and a reason must be given for the need to submit late. Late homework submission permission is not guaranteed.

Study groups are encouraged. However, group answers are not acceptable and each student must submit individual homework solutions. You are encouraged to ask and answer questions on Piazza about the homework assignments, but you should not request for or provide entire solutions. If this behavior is detected, there will be a 50% penalty on your grade for that assignment.

Quizzes
During the first five minutes of each class session, there will be a closed book, closed notes quiz with three to five multiple choice questions about the assigned readings for that week (see textbooks and assigned readings and syllabus schedule). The questions will not involve any mathematical calculation and will assess the student's understanding of some the basic concepts and ideas of the content covered on the assigned readings, which will not necessarily be covered in the lectures. The students are encouraged to ask questions on Piazza about the assigned readings. There will be no make-up quizzes, but we will drop the two lowest quiz scores before calculating the final grade.

Class participation
In preparation for the weekly online meetings, students are expected to watch the lecture videos and read the assigned literature before the start of the meeting. Please be prepared to contribute

to the class discussion: everyone is expected to contribute. In addition, students are encouraged to e-mail questions about the materials covered in the videos and readings of the week to the instructor (raphaeln@umd.edu) before the meetings (deadline for sending questions via e-mail is [Day, Time TBD]).

We will be using Piazza for online out-of-class discussion. The system is highly catered to getting you help fast and efficiently from the instructors and other students. In addition to emailing questions to the instructors for the weekly meetings, we also encourage students to post their questions on Piazza. If you have any problems or feedback for the developers, email team@piazza.com.

Find our class page at: https://piazza.com/piazza.sandbox/fall2017/surv625/home

Participation through Piazza, either by asking or answering questions, is encouraged and positive contributions will be rewarded on your final grade. However, you should not request for or provide entire homework solutions. If this behavior is detected, there will be a 50% penalty on your grade for that assignment.

Final open-book exam
The final cumulative, open-book, take-home exam will be opened on [Date, Time TBD]. Students will have 48 hours to complete it starting from the time the exam is opened on the course website. The solution of the exam should be uploaded to the course website by [Date, Time TBD]. If the student is unable to take the exam on the scheduled week due to prior commitments, he/she should contact the instructor as soon as possible to make special arrangements.

Technical Equipment Needs

The learning experience in this course will mainly rely on the online interaction between students and the instructor during the weekly online meetings. Therefore, we encourage all students in this course to use a web camera and a headset. Decent quality headsets and web cams are available for less than $20 each. We ask students to refrain from using built-in web cams and speakers on their desktops or laptops. We know from our experience in previous online courses that this will reduce the quality of video and audio transmission and therefore will decrease the overall learning experience for all students in the course. In addition, we suggest that students use a wire connection (LAN), if available, when connecting to the online meetings. Wireless connections (WLAN) are usually less stable and might be dropped.

If there are problems or questions with technical issues the student can contact jpsm-itsupport@umd.edu.

Long Course Description

In most statistics courses, students learn numerous data analysis methods and techniques, especially focused on how to make inferences to a population based on stochastic models.

Applied sampling is a unique course in statistics for two reasons: (i) it is concerned almost exclusively with the sample design, and (ii) it adopts a different inferential approach, which does not rely on assumptions about the stochastic distribution of the study variables, but rather on the probability distribution produced by the randomization process on the selection of a probability sample. Therefore, an applied sampling course provides students essential exposure to a different perspective in statistics, which they might not have had contact with in other courses.

As this class is likely the first contact with the subject that most students have had, the primary learning goal we have for students in this course is to understand the basic ideas, concepts and principles underlying sampling methods. The role of randomization in the sample selection and in making inferences to the population, the abstract idea of all possible samples that could have been selected from a population for a given sampling design, and the difference between sampling error and bias are examples of such principles, concepts and ideas.

We also expect that students will acquire some of this basic knowledge through a list of assigned reading we have prepared for this course. Most of them are from the textbook, *Survey Sampling*, by Leslie Kish, which, though written in the 1960s, remains an excellent applied reference. We made this choice because most, if not all, of the more recent books on sampling focus on the statistical theory, but lack a more applied perspective as is presented in Kish. However, many other sampling techniques have been developed since Leslie Kish *Survey Sampling*. For that reason, we also assign more recent research papers about selected topics that are either not covered in the book or that we believe are better developed in other references.

As another learning goal of this course, students should be able to identify and apply appropriate sampling techniques across a range of different methods covered in class (simple random sampling, stratification, cluster sampling, multistage sampling, systematic sampling and probability proportional to size) in sample design problems. As Aristotle once said: "For the things we have to learn before we can do, we learn by doing". We believe this is especially true for statistics: students learn much more by doing exercises. For that reason, we use homework assignments as our main teaching method and assessment for this goal. We carefully design these assignments so students apply what they have learned in lectures and in the readings. Based on the students' performance, we determine which topics need to be reinforced.

Readings

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.

There are also assigned readings of several papers (see list below), available on the course website (jpsmonline.umd.edu):

[1] Rust, K. (1985). Variance estimation for complex estimators in sample surveys. *Journal of Official Statistics*, 1(4), pp. 381-397.

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

[3] Kish, L. and Frankel, M. (1974). Inference from complex samples," *Journal of the Royal Statistical Society*, Series B, **36**, pp. 1 - 37.

[4] American Association for Public Opinion Research (2016). *Standard Definitions: Final Disposition of Case Codes and Outcome Rates for Surveys – Revised 2016*. AAPOR.

[5] Kalton, G. and Kasprzyk, D. (1986). The treatment of missing survey data. *Survey Methodology*, **12**, pp. 1 - 16.

Interested students may find the following recommended books helpful in preparing for the course and as 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).

Academic Conduct

Clear definitions of the forms of academic misconduct, including cheating and plagiarism, as well as information about disciplinary sanctions for academic misconduct may be found at

http://www.graduate.umaryland.edu/policies/misconduct.html
http://www.ugst.umd.edu/courserelatedpolicies.html (University of Maryland) and

https://www.uni-mannheim.de/1/english/research/Good%20Research%20Practice/141119-Satzung%20wiss%20FV%20Senat_en.pdf (University of Mannheim).

Knowledge of these rules is the responsibility of the student and ignorance of them does not excuse misconduct. The student is expected to be familiar with these guidelines before submitting any written work or taking any exams in this course. Lack of familiarity with these rules in no way constitutes an excuse for acts of misconduct. Charges of plagiarism and other forms of academic misconduct will be dealt with very seriously and may result in oral or written reprimands, a lower or failing grade on the assignment, a lower or failing grade for the course, suspension, and/or, in some cases, expulsion from the university.

Accommodations for Students with Disabilities

In order to receive services, students at the University of Maryland must contact the Disability Support Services (DSS) office to register in person for services. Please call the office to set up an

appointment to register with a DSS counselor. Contact the DSS office at 301.314.7682; http://www.counseling.umd.edu/DSS/.

Students at the University of Mannheim should contact the Commissioner and Counsellor for Disabled Students and Students with Chronic Illnesses at http://www.uni-mannheim.de/studienbueros/english/counselling/disabled_persons_and_persons_with_chronic_illnesses/.

Course Evaluation

In an effort to improve the learning experience for students in our online courses, students will be invited to participate in an online course evaluation at the end of the course (in addition to the standard university evaluation survey). Participation is entirely voluntary and highly appreciated.

Class Schedule

Week 1: Introduction; Course Perspectives

Video lecture (Nishimura): available online [Day, Date TBD]
- Design
- Historical
- Typology and vocabulary
- Notation

Readings:
Kish (1965), Chapter 1.

Online meeting (Nishimura): [Day, Date, Time TBD] Includes Quiz

Week 2: Simple Random Sampling, Sampling Frames, and Introduction to Clustering

Homework Assignment 1: due [Day, Date, Time TBD]

Video lecture (Nishimura): available online [Day, Date TBD]
- Simple Random Sampling
- Three Frame Problems
- Frames: Clusters

Readings:
Kish (1965), Sections 2.1-2.7, 11.1-11.3.

Online meeting (Nishimura): [Day, Date, Time TBD] Includes Quiz

Week 3: Stratified Sampling I

Homework Assignment 2: due [Day, Date, Time TBD]

Video lecture (Nishimura): available online [Day, Date TBD]
- Stratified Random Sampling
- Stratified Sampling Allocations

Readings:
Kish (1965), Sections 3.1-3.5.

Online meeting (Nishimura): [Day, Date, Time TBD] Includes Quiz

Week 4: Stratified Sampling II, Systematic selection

Homework Assignment 3: due [Day, Date, Time TBD]

Video lecture (Nishimura): available online [Day, Date TBD]
Stratification Topics
Systematic Selection

Readings:
Kish (1965), Sections 3.6, 4.5A, 4.1-4.3.

Online meeting (Nishimura): [Day, Date, Time TBD] Includes Quiz

Week 5: Cluster Sampling

Homework Assignment 4: due [Day, Date, Time TBD]

Video lecture (Nishimura): available online [Day, Date TBD]
Cluster Sampling
Two-stage Cluster Sampling
Cluster Sampling Subsample Size

Readings:
Kish (1965), Sections 5.1-5.4, 8.3.

Online meeting (Nishimura): [Day, Date, Time TBD] Includes Quiz

Week 6: Unequal-Sized Clusters I

Homework Assignment 5: due [Day, Date, Time TBD]

Video lecture (Nishimura): available online [Day, Date TBD]
Unequal Sized Cluster Sampling
Stratified Unequal Sized Clusters
Complex Designs

Readings:
Kish (1965), Sections 6.1-6.5.

Online meeting (Nishimura): [Day, Date, Time TBD] Includes Quiz

Week 7: Unequal-Sized Clusters II

Homework Assignment 6: due [Day, Date, Time TBD]

Video lecture (Nishimura): available online [Day, Date TBD]
PPS Sampling

PPeS Sampling

Readings:
Kish (1965), Sections 7.1-7.5.

Online meeting (Nishimura): [Day, Date, Time TBD] Includes Quiz

Week 8: Variance Estimation

Homework Assignment 7: due [Day, Date, Time TBD]

Video lecture (Nishimura): available online [Day, Date TBD]
Forming Computing Units
Taylor Series Expansion
Replicated Sampling
Balanced Repeated Replication
Jackknife Replication

Readings:
Kish (1965), Sections 8.6, 4.4.
[1] Rust (1985)
[2] Rust and Rao (1996)
[3] Kish and Frankel (1974)
Kish (1965), Sections 14.1-14.2, 13.1-13.3

Online meeting (Nishimura): [Day, Date, Time TBD] Includes Quiz

Week 9: Final Exam

Homework Assignment 8: due [Day, Date, Time TBD]

Open book, open notes exam.

Note: Student access to the course website will be revoked two weeks after the final exam.