36-303: Sampling, Surveys and Society
Spring 2010

`TR 03:00PM -- 04:20PM BH A53`
`http://www.stat.cmu.edu/~brian/303`
`http://www.cmu.edu/blackboard`

Course Information

Instructor:	*Teaching Assistant:*
Brian Junker, Statistics 132E Baker Hall 268-8874 `brian@stat.cmu.edu`	Jionglin Wu, Statistics FMS 328B (412) 268-1889 `jionglin@stat.cmu.edu`

Office Hours:	*Office Hours:*
Immediately after class *(or by appointment).*	TBA Statistics Tutoring Room FMS Building *(or by appointment).*

Overview

Surveys are ubiquitous in modern life. A recent search for "survey" on Google News yielded 65,122 hits. The first page or so included hits like: "Americans, Amid Recession, Dislike Their Jobs", "Cancer treatment costs play role in doctors' decisions, Tufts survey says", "Survey: Android Gaining Ground, Fast", "Survey: Chile Consumer Price Index Seen Falling 0.3% In Dec", and "A University of Illinois at Chicago research group will perform a survey of River Forest residents to see what they think about police service."

Social and economic policy in the US and elsewhere are informed—and evaluated—by national surveys such as *The National Crime Victimization Survey*, *The National Assessment of Educational Progress*, *The Current Population Survey*, etc. Marketing research regularly uses survey methods; professional journals like *Quirks* regularly have articles like "Searching for representative Black households", "How not to market to Asian-American Youth", "Comparing online and phone research", "Keep undesirable respondents out of your online survey". And survey sampling is an indispensable part of economics, political science, and other social sciences.

Yet survey methodology as we know it today was really only developed in the 20^{th} century, and much of the growth in the use of surveys has occurred in just the past 50 years. At the beginning of the 20^{th} century statisticians were debating whether anything less than a complete census—asking *every* member of a target population—could be reliable, yet today, samples of 500 or 1,000 individuals from populations of millions of people are routinely used. A reliable survey has many interrelated components: identifying a target population, choosing a sampling method, deciding how to collect data from each survey respondent, writing good survey questions, dealing with missing responses, analyzing the data, making a report.

In this course we will look at the role of sampling and sample surveys in the context of U.S. society and its institutions. We will examine the evolution of survey taking in the United States in the context of its economic, social and political uses. This will eventually lead to discussions about the accuracy and relevance of survey responses, especially in light of various kinds of nonsampling error. All students will be required to design, implement, analyze, and report on a sample survey.

Textbooks, Resources and Statistical Methods

The required text (in the bookstore) for the course is:

- Groves, R. M., Fowler, F. J. Jr., Couper, M. P., Lepkowski, J. M., Singer, E. and Tourangeau, R. (2004). *Survey Methodology* (paperback). New York: Wiley. ISBN: 978-0471483489. ($58.10 at Amazon.com)

Some other useful resources on survey sampling include (these are not in the bookstore):

- Converse, J. M. & Presser, S. (1986). *Survey Questions: Handcrafting the Standardized Questionnaire* (paperback). Sage Publications Inc. ISBN: 978-0803927438. ($14.42 at Amazon.com)
- Kalton, G. (1983). *Introduction to Survey Sampling* (paperback). Sage Publications, Inc. ISBN: 978-0803921269. ($16.61 at Amazon.com)

and if you are really interested in survey sampling, the very next serious book you are likely to read is:

- Lohr, S. (1999). *Sampling: Design And Analysis.* (hardcover). Duxbury Press. ISBN: 978-0534353612. ($143.00 at Amazon.com)

There will also be a variety of handouts, examples, and lecture notes at the course website.

Some computation and statistical analysis will be required in the course. You are welcome to use any statistics package you are familiar with, such such as Excel, Minitab, R, Splus, or SAS (all are available in public clusters; for your personal computer, Minitab can be downloaded from `www.cmu.edu/myandrew` and R can be downloaded from `http://cran.r-project.org/`).

You will be using statistical methods you have learned in 36-202 or a similar course. *There will be a quiz in the second lecture covering the material you need for this course.*

Objectives

1. To gain an appreciation for the uses of sample surveys in American society;
2. To learn the statistical properties associated with sampling from a finite population ;
3. To learn about practical aspects of the design and implementation of sample surveys;
4. To understand the nature of non-sampling and measurement errors in surveys and how to control their effects;
5. To design, conduct, analyze, and report on a sample survey.

The course will run in two parallel "tracks": The "Classwork" track will give you some of the tools you need to do the "Project" track of the course. *Both tracks* of the class will start right away, to give you enough tools and enough time to successfully complete your projects.

Requirements: Classwork

Classwork will consist of the following four requirements:

- Reading assignments from the textbook and other sources, and taking notes in class;
- Quizzes and classroom discussion;
- Homework assignments (practice with writing, computation, and other tasks);
- Two midterm exams.

Clear, careful writing and interpretation of results is an important part of both weekly homeworks and the projects.

Always be judicious about including computer output and graphs: show enough that we can clearly see what you are doing, but not so much that we will get lost or bored leafing through your work! A good rule of thumb is to remove any figures, tables, graphs, etc. that you do not have something interesting to say about in the text of your homework solution.

Some homework will be submitted to me in class, and some will be submitted electronically by email or on Blackboard. For electronic submissions, *only one pdf or text file will be accepted, per assignment.* Please do not submit ms word documents; convert them to pdf and submit that instead. Because of font and version differences, ms word may look fine on one computer but be unintelligible on another.

Requirements: Project

For the project you will divide up into groups of about 4–5 students each. Each group will select a survey topic, design questions and a sampling plan, carry out the survey, analyze the results, and produce a final written report.

Clear, careful writing and interpretation of results is an important part of both weekly homeworks and the projects.

Requirements: Grading

The various requirements in the course will be weighted as follows:

Homework, Quizzes, Other Assigments	20%
Midterm (FEB 18)	15%
Final (APR 13)	15%
Citizenship, Discussion, Attendance	10%
Project Phase I	5%
Project Phase II	5%
Project Phase III	30%
Total	100%

Extra Credit

If your group registers for Meeting of the Minds (MoM; Deadline Apr 1) and presents a poster (May 5), I will throw out the lowest exam score of each person in that group. If your group registers for the Statistics Poster Competition at MoM and wins first place, I will give that group all A's regardless of their numerical scores. In addition, first place always gets a cash prize from the Statistics Department.

Citizenship

Much of the work done for this course relies on working with others. You are expected to participate fully in both class and in group assignments. You will be graded on this by both the instructor and the other members of your group. Occasionally I will ask you to fill out peer citizenship assessment forms for your group. These evaluations are confidential and anonymous in the sense that only I and the TA will see the results; other students in the class will not. I will, however, provide feedback if necessary. The purpose of the assessment is only to ensure that everyone is participating fully and working together, NOT to assess someones strengths and weaknesses with the class material. Your citizenship grade will be worth 10%.

Plagiarism

Plagiarism is completely unacceptable. For the purposes of this class, plagiarism is using other peoples words or ideas without crediting those people.

- Group work can use the words and ideas generated by that group without giving credit other than listing all group members names on the assignment.
- If you work with someone on an assignment, simply acknowledge that you did so on the first page.
- If you get ideas or words from a website, journal article, etc that you found helpful, cite the source on your assignment/group work, right where you use it. Then put a bibliography or list of sources cited at the end of each piece of work you do.
- You will not be penalized for working together (except for exams) if appropriately cited.
- Each case of plagiarism will be considered individually; consequences will depend on the seriousness of the case.

Rough Plan of Action

Following is a **rough** outline of the topics we will cover, together with associated chapters from Groves et al. *Additional readings may also be assigned.* We are reading the chapters of Groves out of order, so that you can start sooner on some important project tasks.

A more detailed lecture plan will be updated weekly on `http://www.stat.cmu.edu/~brian/303`.

Week: Topic	Date/Reading	Project Milestone
WEEK 01: Introduction	*Jan 12 / Jan 14*	I.0: Group Requests
History, Examples	Groves, Ch 1	
Key elements of samples	Groves, Ch 2	
DIAGNOSTIC QUIZ THURSDAY JAN 14		
WEEK 02: Inference, Error and Ethics	*Jan 19 / Jan 21*	I.0: Group Assignments
Target Population, Sampling Frames, Coverage Error	Groves, Ch 3	
Ethics/IRB	Groves, Ch 11	
WEEK 03: Designing Survey Questions	*Jan 26 / Jan 28*	I.1: Two Topic Proposals Per Group
Methods of Data Collection	Groves, Ch 5	
Sampling from a Finite Population	Groves, Ch 4	
	Kalton, Lohr Excerpts	
WEEK 04: Evaluating Survey Questions	*Feb 2 / Feb 4*	I.2: Population, Frame, Mode (for BOTH topics!)
Sampling from a Finite Population	Groves, Ch 4	
	Kalton, Lohr Excerpts	
Questions and Answers in Surveys	Groves, Ch 7	
	Converse/Presser Excerpts	
WEEK 05: Catchup & Case Studies	*Feb 09 / Feb 11*	I.3: Pick Topic, Start IRB
Evaluating Survey Questions	Groves, Ch 8, 9	
	Converse/Presser Excerpts	
WEEK 06: Project Progress Reports & Midterm EXAM	*Feb 16 / Feb 18*	II.4: Sampling Scheme & Design Questions
MIDTERM EXAM THURDSAY FEB 18		
WEEK 07: Sampling from a Finite Population	*Feb 23 / Feb 25*	II.5: Pretest Questions
Unit and Item Nonresponse	Groves, Ch 6	
Sensitive Questions	TBA	
WEEK 08: Catch up and Case Studies	*Mar 2 / Mar 4*	II.6: Final IRB and Project Plan
MIDSEMESTER BREAK	*FRI MAR 5*	
SPRING BREAK	*MAR 08 – MAR 12*	
WEEK 09: Topics, e.g. Stratified Sampling	*Mar 16 / Mar 18*	
	Kalton, Lohr Excerpts	
WEEK 10: Topics, e.g. Clustered Sampling	*Mar 23 / Mar 25*	
	Kalton, Lohr Excerpts	
WEEK 11: Preparing Survey Data for Analysis	*Mar 30 / Apr 1*	III.7: Progress Reports
MEETING OF THE MINDS DEADLINE APRIL 1	Groves, Ch 10	
WEEK 12: Catchup & Case Studies	*Apr 6 / Apr 8*	III.8: Rough Drafts
	Groves, Ch 13	
WEEK 13: Final EXAM & Carnival	*Apr 13 / Apr 15*	
FINAL EXAM TUESDAY APRIL 13		
WEEK 14: Special Topics, Project Presentations	*Apr 20 / Apr 22*	III.9: Oral Presentations & Final Reports
WEEK 15: Special Topics, Project Presentations	*Apr 27 / Apr 29*	
MEETING OF THE MINDS	*May 05*	