

Survey Sampling (MATH-658)

Georgetown University
Summer 2015

Syllabus

Time and Room: Mon Wed 6:00 - 8:30 pm, June 1 through July 24, in STM, Room *TBA*.

Instructor: Tommy Wright; (301)214-2262 (home); (301) 763-1702 (office)
Email: tw44@georgetown.edu; tommy.wright@census.gov (Only read in evenings)

Office Hours: 4:30 - 5:45 pm before each class in Room STM *TBA*. Immediately following class. Other times by appointment.

Expectations:

- 100 percent attendance is expected of each student. Please notify instructor when unable to attend.
- The goal is to understand concepts, definitions, and results as they come.
- Individual work is expected on Exercises(Homework); Examination I; and Examination II.

Prerequisites: Elementary probability and knowledge of the Central Limit Theorem.

Objectives:

- To study and understand the concepts underlying modern survey sampling methodology and its usefulness as a tool for collection of data for effective decision making.
- To apply some of the learned methodology through execution of a small scale data collection project as part of a team effort. (SAMPLING PROJECT: Need Statement, Target Population, Sampling Frame, Primary Parameter(s), Questionnaire/Data Collection Instrument, Sampling Plan, Estimation Plan, Pilot, Sample Size, Report, and Presentation.)

Background: “All scientific observation, whether statistical or not, is based on *sampling*,” says Stephan(1948). “The earliest examples of sampling procedures are to be found in certain very ordinary human activities. The common practice of taking a small part or portion for tasting or testing to determine the characteristics of the whole precedes recorded history and is one of the roots from which sampling methodology stems...”

The current approach to sampling assumes a given finite collection of units, called a *population* (or *universe*). When examination of each and every unit in the population is undesirable to know a particular population characteristic or *parameter*, a *sample*, i.e., a subset or portion of the population, may be selected to yield satisfactory information regarding the particular population characteristic. The population characteristic is often a quantitative one. In such cases, a *statistic* is computed using information collected from the small and more manageable sample, and its value is used to *estimate* the unknown value of the population characteristic. Although we desire a sample that will provide a “good” estimate of the unknown value of the population characteristic, it is certainly conceivable that the sample information obtained could lead to a very incorrect estimate.

Probability sampling makes use of the laws of probability in the selection of the sample and in the construction of efficient estimators. With *probability sampling*, every population unit has a known positive chance of being selected for the sample. Probability sampling provides a means for saying how good one believes an estimate is relative to all the possible estimates from all of the possible samples. Making use of probability according to the way in which the sample is selected by use of randomization theory is called the *design-based approach*. When probability is induced by making use of models which provide probability distributions, we refer to this as the *model-based approach*. We

will make primary use of the conventional design-based approach to survey sampling with some references to the model-based approach. This combined approach is often referred to as the *model-assisted approach*.

Business applications of sampling techniques are numerous for decision making and include: marketing surveys, auditing, customer surveys, economic surveys, accounting, and product quality control.

In fact, sampling methods are used throughout the world by a variety of individuals, groups, and organizations, as well as by local, state, and national governments. They are used successfully in many other fields including agriculture, defense, economics, education, energy, environment, finance, health, industry, labor, natural resource management, demographics, social and behavioral sciences, and transportation. Some specific applications include taking opinion polls, election polls, and polls for rating television programs, surveying animal populations (particularly fish, deer, etc) and farms; taking air samples to monitor air quality; sampling to monitor traffic activity; sampling to estimate energy consumption; sampling to monitor a nation's economy; sampling to monitor employment; sampling to monitor education progress; and taking samples of products produced at a manufacturing plant to monitor output quality. The complexity of sampling methods can vary greatly from one application to another.

Textbook: Sharon Lohr, *Sampling Design and Analysis*(2010), Brooks/Cole (Cengage learning) Publishing Company, Boston, MA: ISBN 13-978-0-495-10527-5.

Topics Include: Chapters 1, 2, 3, 4, 5, 6, and selected sections of chapters 7, 8, 9, 10, 11, and 12. Also, Appendix A. We will consider selected topics in probability, simple probability sampling (e.g., simple random sampling, systematic sampling), ratio and regression estimation, stratified sampling, cluster sampling, complex sampling, nonresponse, analysis of data from surveys, etc.

Homework: Homework will be assigned regularly, collected, and graded. Homework will consist of *exercises* and *problems* (assigned at one class meeting and due at the beginning of the next class meeting). Only a sample of the exercises/problems will be graded each time, so it is important to do all work carefully. Late or missing homework assignments will be assigned a score = 0. The three lowest homework assignments will be dropped before the average is computed.

Examinations:

EXAMINATION I (In-class: Tentatively during the 1st hour on Wednesday, June 24).

EXAMINATION II (Take-home: TBA).

Project Teams:

Team (A, B, C, D) Formations (Last 30 minutes on June 8).

Team Meetings (Last 30 minutes on various days; Other times).

Team Reports/Presentations (Each team on July 24).

Evaluation:

- (i) Homework Average 33 points
- (ii) Examination I (In-class) 33 points
- (iii) Examination II (Take-home) 33 points
- (iv) Team Project Reports/Presentations 34 points

Any Combination of the top two totals among (i), (ii), and (iii) **plus** (iv) = 100 points.

Grading:	A	95 points or more	B	80 points or more	F	Below 70 points
	A-	90 points or more	B-	75 points or more		
	B+	85 points or more	C	70 points or more		