Math 315 Introduction to Sampling Techniques

Professor: Katie St. Clair

Contact Info: kstclair@carleton.edu, CMC 220, 507.222.4193

Office Hours: TBD

Class Web: `https://moodle.carleton.edu`

Textbook: *Sampling: Design and Analysis* 2ed, Lohr

Technology: We will use the statistical software R which is available in the statistics lab (CMC 201) and is free to download at `www.r-project.org`. You will need a calculator for exams and some assignments.

Description: We will cover a wide range of statistical sampling techniques that are used to make inferences about a population. We will start with the most basic of designs: a simple random sample, then quickly add complexity with stratification, clustering, and unequal selection probabilities. We then discuss how to form estimates of unknown population parameters and quantify our sampling error when data is collected using a complex sampling design. By the end of the term you will know how to graph, run chi-square tests, and fit basic regressions models for complex survey data. The course will cover sampling techniques used in the study of both human populations and natural resources, and we will use the R "survey" package extensively throughout the course.

Goals: The primary learning goals for this course are

- understand the objectives of a sample survey
- know the "common" sampling designs, recognize the design in an application, and understand when it may be appropriate to use each design.
- understand how we can compare sampling designs or estimators, and know what properties make a "better" design or estimator.
- analyze survey data using the statistical software R
- design and implement your own survey
- understand the sampling methodology used in many research and government surveys
- present survey analysis results in written and oral formats

Class Expectations: Come to class prepared - read assigned material and bring your book and class notes/handouts, turn off cell phones, and be on-time to class.

Grading: Your final grade will be a weighted average of the following:

5% - Participation and Attendance: This grade is based on your group participation scores, classroom behavior, involvement in discussion and activities, and attendance (no more than 2 absences).

12% - Homework: Homework will be assigned weekly. You may work with classmates but you must write up your assignment on your own. Always follow the **homework guidelines** given on the class web page.

5% - Short report: You and a partner will design and implement a sampling plan to estimate the size of a given crowd, analyze the data, construct a size estimate, and report your results in two-page paper. The crowd in question will be one of my making, but it is designed to mimic what might be seen using a ariel image of a large crowd.

15% - Group Project: Your group will select a population to study, design and implement a sampling strategy, analyze the data to address specific questions of interest, and present your results in a five-page report (10%) and brief oral presentation (5%). I will assign groups for this project and need to approve your population before you proceed with the project. One very large restriction for this project: no human populations allowed! You will also be graded according to the complexity of your design or analysis, so try to avoid using only a SRS.

48% - Exams 1 and 2: There will be two midterms during the term. Each exam is worth 24%.

15% - Final Paper: You will find a data set online that was collected using a complex sampling design, analyze the data using methods discussed in class, and present your results in a paper. This will be an individual paper done in lieu of a final exam.

Course Content: The major course topics are as follows:

Intro survey design vs. sampling design, sources of bias, sources of error, parameter vs. statistic

Sampling designs SRS, stratification, clustering (one and two-stage), comparing designs using computer simulations, sampling with unequal probabilities, two-phase sampling, capture-recapture, multistage sampling

Estimation how sampling theory differs from "stat 101" theory, how do statisticians derive parameter estimates and their standard errors, using sampling weights in estimation, using the R "survey" package, confidence intervals, comparing estimators, comparing designs using design effects, improving estimates using auxiliary information (ratio and regression estimators)

Additional topics graphing complex survey data, regression analysis and chi-square tests for complex survey data, nonresponse issues