Analysis of Complex Survey Data

CLASS SESSIONS
Friday, June 28, 2019
8:30am - 5:30pm
(Includes 1 hour lunch break)
Location: TBD
Directions can be found here: http://www.cuepissummer.org/contactpage

INSTRUCTORS
Natalie S Levy, MPH, PhD Candidate
Email: nsl2110@cumc.columbia.edu
Office location: 722 W 168th St., 5th floor, Room 511

Luis E Segura, MD, MPH, DrPH Candidate
Email: les2196@cumc.columbia.edu
Office location: 722 W 168th St., 5th floor, Room 511

COURSE DESCRIPTION
Complex survey data violate typical assumptions about simple random samples of independent observations, thus requiring specialized statistical techniques. This course will provide participants with practical skills to analyze data arising from complex epidemiologic sampling designs. The theory behind complex sampling strategies and the necessity of applying appropriate statistical techniques to analyze these data and make valid inferences will be discussed. National Household Survey on Drug Use and Health (NSDUH) data will be used for applied demonstrations, illustrating concepts applicable to all datasets arising from complex survey designs. We will demonstrate the appropriate use of sampling weights in the NSDUH data and how the appropriate weight is specific to the research question being asked. We will demonstrate how to obtain basic descriptive statistics, appropriate variance estimates, and regression parameters in R (SAS and Stata code will also be provided).

PREREQUISITES
None.

COURSE LEARNING OBJECTIVES
Students who successfully complete this course will be able to:

- Understand the theoretical basis for complex survey designs and its relationship to external validity.
- Understand the influence of design effects on standard errors and why special statistical procedures are needed to analyze complex survey data.
- Have familiarity with the National Surveys on Drug Use and Health.
- Use R software to analyze complex survey data including: basic descriptive analysis, bivariate analysis, and regression models (SAS and Stata code will also be provided).

REQUIRED READINGS
3. Survey Analysis in R. http://r-survey.r-forge.r-project.org/survey/

RECOMMENDED READINGS

COURSE STRUCTURE
Class time is 8 hours total. The structure of the workshop will include a mixture of lecture and guided exercise with hands-on assistance from the instructors.