The University of North Carolina at Chapel Hill

Spring Semester 2014

BIOS 613-001

Working with Data in a Public Health Research Setting

I. Time and place
   TuTh 3:30-4:45 PM, McGavran-Greenberg 2308

II. Instructor
    Kathy Roggenkamp
    Research Instructor and Manager of Statistical Computing
    Dept. of Biostatistics and Collaborative Studies Coordinating Center
    Office: Room 206, Suite 203, CVS Plaza, 137 E. Franklin St.
    Phone: 919-966-5304
    E-mail: kathy Roggenkamp@unc.edu
    Office hours: By arrangement (after class will generally work)

III. Textbook
     None

IV. Course description

   This three-credit course has a target audience of MS or MPH-seeking biostatistics students who are in their second semester of study. It aims to provide a conceptual foundation and practical training to these students in various aspects of working with data, since they will be using data from clinical trials or other public health research studies while in graduate school and after graduation. Topics include using SAS and SQL to transform data into structures useful for analysis, producing typical reports, working toward study data of high quality, working with big data, and simulation with SAS. Prerequisite: BIOS 511, EPID 700, or permission of the instructor (basically, solid knowledge of SAS DATA step programming and familiarity with the SAS macro facility)

V. Course format

   • Students will be expected to prepare for class by reviewing materials as specified by the instructor in the detailed course schedule. Such materials could include articles, instructor-provided notes, or prepared videos. Materials will be provided on Sakai.
   • Since students will have reviewed relevant material before class, most class time will be spent working on problems, often as part of a small group.
   • To perform these activities, students will be required to bring a laptop or other portable device to class. Students will be encouraged to use a BIOS 613-specific virtual session (VCL) for running SAS and other software, or they can use software loaded locally on their device.
• Students are required to attend class and participate in class activities.
• Daily assignments will generally be turned in (electronically via Sakai) and graded on a three-level scale: $V^+, V, \text{ or } V^-\).
• Overall, the course will be graded as follows: H (high pass)=exceeds expectations, P (pass)=meets expectations, L (low pass)=fails to meet expectations. I expect that most students will receive a P. Since no undergraduates are enrolled in the course, this grading scheme will work.
• There will be no written exams during the run of the course.
• The course’s final exam is officially scheduled for April 29, 2014.
• Near the end of the semester, students will work on a course-related project of their choosing and present their work to the class.
• Since this will be the initial offering of a new incarnation of BIOS 613, the spring 2014 course should be considered a first draft. The instructor will provide students with changes in plans as early as she reasonably can.
• In future years the course number will be BIOS 669.
• Comments and contributions that will enhance the course in future years will be greatly appreciated.

VI. List of topics – most topics will cover multiple class sessions

• SAS refresher and an introduction to the METS clinical trial (METS data will be used for many course exercises)
• PROC SQL, including an introduction to relational databases
• Look-up tables
• Analysis data sets and variables, including an NHANES exercise
• PROC REPORT and general reporting concepts
• Data quality, covering data transitions, missing values, and data cleaning
• Exercises using Orange County Health Department data
• Metadata
• Data closure and codebooks
• Big data (only an introduction, but will include an exercise)
• Simulation in SAS, using both base SAS and IML (only an introduction)