SCHOOL OF MEDICINE & DENTISTRY
UNIVERSITY of ROCHESTER MEDICAL CENTER

Syllabus

BST 461 – Biostatistical Methods I
Fall, 2021

Instructors:
Tong Tong Wu, Ph.D.
Phone / E-mail: 276-6858 / Tongtong_Wu@urmc.rochester.edu

Michael P. McDermott, Ph.D.
Phone / E-mail: 275-6685 / Michael_McDermott@urmc.rochester.edu

Teaching Assistants: XXX
E-mail: XXX

Office Hours: By appointment

Classes: Monday / Wednesday, 1:30-3:10 PM, SRB 1.406

Dates: August 25 – December 8

Course Description
This is a course on statistical methods commonly employed in the analysis of data in biomedical research. The course will focus on application of statistical methods rather than their underlying theory. Inferential procedures (hypothesis testing, interval estimation), checking of model assumptions, model building and refinement, and interpretation of results will be emphasized. Some elements of study design, including sample size determination, will be introduced. SAS software will be used to illustrate the methods. More advanced courses (BST 426 – Linear Models; BST 479 – Generalized Linear Models) are devoted to the theory underlying most of the methods covered in this course.

The following is a list of the major topics to be covered in the course (approximate timing in brackets):

- *Elements of study design*: Cross-sectional, cohort, and case-control studies; randomized controlled trials; use of devices such as blocking, stratification, and matching [Week 1; Professor McDermott]
- *Categorical data overview*: Categorical response data; discrete distributions; Wald, likelihood ratio, and score tests [Week 2; Professor Wu]
- *Analysis of screening tests*: Sensitivity and specificity; positive and negative predictive values; ROC curves [Week 3; Professor Wu]
- *Two-way contingency table analysis*: Methods for comparing proportions; measures of association; chi-square tests; inference for ordinal tables (linear and monotone trend); Fisher's exact test [Weeks 4-6]; Professor Wu]

- *Three-way contingency table analysis*: Conditional vs. marginal tables (odds ratios, independence); Simpson's paradox; Cochran-Mantel-Haenszel test; assessing homogeneity of associations in 2 x 2 x K tables [Weeks 6-7; Professor Wu]
- *Analysis of paired categorical data*: McNemar's test; kappa and weighted kappa; Bradley-Terry model [Weeks 7-8; Professor Wu]
- *Linear models*: Analysis of variance; regression and correlation; multiple regression; analysis of covariance; confounding; interaction [Weeks 8-12; Professor McDermott]
- *Sample size determination* [Weeks 12-13; Professor McDermott]
- *Generalized linear models*: Components of generalized linear models; logistic regression [Weeks 13-14; Professor Wu]
- *Presentation of student projects* [Weeks 14-15]

Prerequisites
An introductory course in statistics (similar to BST 463 – Introduction to Biostatistics).

Course Aims and Objectives
The course has five major objectives:

1. To provide an overview of various aspects of sound study design.
2. To provide a detailed understanding of the application of statistical methods to yield inferences about population parameters that address the scientific questions / hypotheses of interest.
3. To illustrate the use of descriptive tools to examine the appropriateness of the assumptions underlying the statistical methods that are used.
4. To provide an overview of methods used to build and refine statistical models.
5. To provide an introduction to the use of SAS software to appropriately apply the methods covered in the course.

Course Policies and Expectations
The following policies/expectations apply to this course:

- You are expected to attend all classes. There is a required component of class participation and questions are encouraged.
- You should have your video turned on during all Zoom sessions.
- Cell phones should be turned off or set to vibrate for emergencies.
- We do not have a fixed schedule for office hours, but we are more than happy to make appointments with students via Zoom or in-person to address questions or anything else related to the class.
- You are encouraged to discuss homework assignments with other students in the class, but you must write up the final solutions yourself. Please refer to the department policy on student collaboration: https://www.urmc.rochester.edu/biostat/courses/studentcollaboration.aspx
- Late homework assignments will be permitted only in case of emergency or with prior approval of the instructor.

Course Materials

The primary material for the course will be the presentation slides and output from SAS programs. The following books may also be useful:

1. Agresti A. An Introduction to Categorical Data Analysis, Second Edition. Hoboken, NJ: John Wiley and Sons, 2007. This is available in the department library (CAT 5.705) and through the University of Rochester Library as a free eBook (https://ebookcentral.proquest.com/lib/rochester/detail.action?docID=290465).
2. Pepe MS. The Statistical Evaluation of Medical Tests for Classification and Prediction. Oxford, UK: Oxford University Press, 2003. This is available in the department library (EPI 373.000, EPI 373.001, EPI 373.002) and in Carlson Library (RB38.3.P47 2003).
3. Rosner B. Fundamentals of Biostatistics, Sixth Edition. Belmont, CA: Thomson Brooks/Cole, 2006. This is available in the department library (TEXT 380.102).
4. Kutner MH, Nachtsheim CJ, Neter J, Li W. Applied Linear Regression Models, Fifth Edition. New York, NY: McGraw-Hill/Irwin, 2005. This is available in the department library (LINM 273.000, LINM 273.001).
5. Hosmer DW Jr, Lemeshow S, Sturdivant RX. Applied Logistic Regression, Third Edition. Hoboken, NJ: John Wiley and Sons, 2013. This is available in the department library (APPL 211.012) and through the University of Rochester Library as a free eBook (https://ebookcentral.proquest.com/lib/rochester/detail.action?docID=1138225 or https://www.oreilly.com/library/view/applied-logistic-regression/9781118548356/?ar).

Material for the course will be drawn from all of these sources (and others).

The homework assignments will involve the use of SAS software. There will be in-class instruction on how to write SAS programs to analyze data using the methods introduced in the course. The following resources should be helpful:

- Delwiche LD, Slaughter SJ. The Little SAS Book: A Primer, Sixth Edition. Cary, NC: The SAS Institute, Inc., 2019. This is available through the University of Rochester Library as a free eBook (https://library-books24x7-com.ezp.lib.rochester.edu/toc.aspx?site=WLKOY&bookid=148247 or https://www.oreilly.com/library/view/the-little-sas/9781642953435/?ar/).

- SAS documentation
  https://documentation.sas.com/doc/en/pgmsascdc/9.4_3.5/pgmsassyntaxwlcm/home.htm

- Statistical computing from UCLA
  https://stats.idre.ucla.edu/r/
  https://stats.idre.ucla.edu/other/examples/icda/

Assignments and Grading Procedures
The final grade will be based on homework assignments (60%), class participation (10%), and a final project (30%). In general, homework assignments will be due 1-2 weeks after being distributed. Graded assignments will be returned within one week.

Each student needs to complete a project involving the analysis of a data set of his/her choosing. One of the instructors will be assigned to supervise the student on the project. The student will meet with the assigned instructor to discuss the data set, research question, and analytic methods to be used. The project has to be approved by the instructor. Methods and techniques learned in this course should be applied to analyze the chosen data set. Students will give 20-minute presentations of their projects during the final two classes of the semester. A written report must be submitted during the final examination period and should include the following:

Introduction: A description of the background to the project should be provided, including a brief account of the scientific rationale and objectives of the study.

Methods: Some topics to consider are type of study (observational or experimental); study population; method of sampling and choice of sample size; description of intervention(s), if any; method of randomization, if any; response variables; and covariates and potential confounding variables. A description of the statistical methodology used should be provided, with references to the literature where appropriate. The description should be detailed enough to allow another competent statistician, given the same data and computing facilities, to reproduce major steps in the analysis.

Results: A summary of the major results from the analysis should be provided using both descriptive and inferential methods. The assumptions of any statistical models used should be checked using appropriate methods; the results of model verification should also be reported.

Conclusions: A clear interpretation of the findings of the study should be provided. This section must be written for general readers and should be in non-statistical language.

Academic Integrity
Academic integrity is a core value of the University of Rochester. Students who violate the University of Rochester "University Policy on Academic Honesty" are subject to disciplinary penalties, including the possibility of failure in the course and/or dismissal from the University. Since academic dishonesty harms the individual, other students, and the integrity of the University, policies on academic dishonesty are strictly enforced. For further information on the University of Rochester Policy on Academic Honesty, please visit the following website:
http://www.rochester.edu/college/honesty/docs/Academic_Honesty.pdf

Accommodations for Students with Disabilities
Students needing academic adjustments or accommodations because of a documented disability must contact the Disability Resource Coordinator for the school in which they are enrolled (see link below for contact information).
http://www.rochester.edu/eoc/DisabilityCoordinators.html

Course Schedule
Classes will be held twice per week, but please note that there will be no class held on September 6 (Labor Day) or November 24 (Thanksgiving Break).