Biostatistical Methods: Survival Analysis and Causality
PB HLTH C240B / STAT C245B
Fall 2017

Registration Information: Public Health C240B | 4 Units | Class Numbers 45292 & 45293
Statistics C245B | 4 Units | Class Numbers 67300 & 67301

Time & Location: TuTh 2:00P-3:30P | 2030 Valley Life Sciences
W 11:00A-11:59A | 342 Evans

Instructor: Mark van der Laan
E-mail: laan@berkeley.edu
Website: www.stat.berkeley.edu/~laan
Office Hours: Tu 4:00-5:00P | Haviland 108
or by appointment

GSI: Jonathan Levy
E-mail: jlstiles@berkeley.edu
Office Hours: W 2:00-3:00P | Haviland 111
or by appointment

Course Description

Analysis of survival time data using parametric and non-parametric models, hypothesis testing, and methods for analyzing censored (partially observed) data with covariates. Topics include marginal estimation of a survival function, estimation of a multiplicative intensity model (such as Cox proportional hazards model) and estimation of causal parameters such as treatment specific survival functions and contrasts thereof. Additionally, general theory for developing efficient estimators of the parameters of interest in censored longitudinal data models, including empirical process theory, influence curve, functional delta method, efficient influence curve, efficiency theory, targeted maximum likelihood (TMLE) estimation, and general proofs of efficiency of MLE and TMLE. Computing techniques, numerical methods, simulation and general implementation of biostatistical analysis techniques with emphasis on data applications.

Instructional Strategy

Most pedagogical studies (i.e. those concerned with the methods and effectiveness of teaching) indicate that lectures by themselves are a poor way of engaging students and promoting learning. To address this problem, this course will use a Flipped Classroom model. This involves shifting the material presented in class and out of class. Instructional core content is delivered online, outside of the classroom. Class time is spent exploring topics in greater depth and creates meaningful learning opportunities. This rearrangement allows for more interactive, active learning opportunities during class time like group discussion, Q&A, problem solving activities, and R labs where students will apply the methods presented to real data. It also allows for self-paced comprehension of highly complex core concepts. Video lectures give students the ability to pause, rewind, and even re-watch content delivery opposed to traditional lectures that require content delivery to occur in a fixed time and place.

The video lectures are intended to serve as jumping off points to drive discussion, activities, and clarification during class time. Thus, it is essential that students watch the video lecture and take notes **before** class. Coming to class confused is welcome, coming to class empty-headed is ineffective under this model.

Biostatistical Methods: Survival Analysis and Causality
PB HLTH C240B / STAT C245B
Fall 2017

Learning Outcomes

At the end of this course students should be able to:

- Utilize the R statistical software in order to:
    - Apply both classical and innovative survival methods/estimators like Kaplan-Meier, Cox proportional hazards, multiplicative intensity models, TMLE of treatment specific survival curves based on longitudinal data, EM algorithm to compute NPMLE for censored data
    - Construct and interpret influence curve based confidence intervals including simultaneous confidence intervals
    - Design simulations for the practical evaluation of these estimators and inference
- Derive an influence curve of an estimator with the functional delta method
- Derive an efficient influence curve/canonical gradient of a path-wise differentiable target parameter by projecting onto the tangent space
- Discuss/explain asymptotic linearity, the basics of empirical process theory, and the functional Central Limit Theorem for empirical processes
- Apply the roadmap of statistical learning to formulate an estimation problem in a survival setting

Grading Policy

Assignments – 40%

Encompass both in-class and out-of-class work compromising theory and data analysis using R. Assignments include homework problems and R labs. There will be about five homework assignments and about five R labs.

Final Group Project – 30%

Consists of a written report and a presentation on a topic that involves the application of survival analysis statistical methods and software to address a particular question of interest. Details will be provided later in the semester.

Participation – 20%

Includes attendance, engagement during class, listening to lectures before class, team-based learning exercises, blog involvement, and course evaluations.

Scribing – 10%

Involves the collaborative scribing of video lectures to develop a collection of high quality notes for students to keep. Every student will be required to scribe some of the lectures in LaTeX. Once the scribed notes are reviewed by the GSI, they will be available to the class on bCourses. A sign-up sheet and more details will be provided on bCourses.

Biostatistical Methods: Survival Analysis and Causality
PB HLTH C240B / STAT C245B
Fall 2017

Materials

The weekly updated schedule, lecture videos, scribed notes, R lab assignments, homework assignments, supplemental materials, and blog information will be provided on bCourses.

The following textbooks serve as excellent references, encompassing all of the topics covered in this course:

- *Targeted Learning: Causal Inference for Observational and Experimental Data* by Mark van der Laan and Sherri Rose (2011)
- *The Statistical Analysis of Failure Time Data, Second Edition* by John Kalbfleisch and Ross Prentice (2002)

Prerequisites

STAT 201A-B (may be taken concurrently) or older version STAT 200A-B or consent of instructor. PB HLTH C240A / STAT C245A is recommended. Familiarity with the R language.