University of California, Berkeley
Fall 2019 Syllabus

Biostatistical Methods: Survival Analysis and Causality

Registration Information: Public Health C240B | Course Control Numbers 32592 & 32593
Statistics C245B | Course Control Numbers 67300 & 67301

Time & Location: TuTh 12:30P-1:59P | 330 Evans

Instructor: Mark van der Laan
E-mail: laan@berkeley.edu
Office Hour: Th 2:00P-3:00P | 5311 BWW

GSI: Rachael Phillips
E-mail: rachaelvphillips@berkeley.edu
Office Hour: Tu 11:30A-12:30P | 330 Evans

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 *Blended Learning/Hybrid Classroom* format. This involves shifting the majority of the material presented in class to 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 labs where students will apply the methods presented to real data. It also allows for self-paced comprehension of 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 recommended that students watch the video lecture *before* class. Under this instructional model, coming to class confused is welcome, but coming to class empty-headed is ineffective.

Prerequisites

It is highly recommended for students to have some higher-level background knowledge in statistics and mathematics (e.g., STAT 201A-B, STAT 200A-B, PB HLTH C240A / STAT C245A). Also recommended is data analysis experience and/or familiarity with software in the R programming language.

Note: We welcome students with diverse backgrounds to this course. If a student does not have the recommended background knowledge, then the course may be more challenging but they are not guaranteed to fail. Students who attend, participate, complete assignments on time and try their best will succeed in this course.

Credit Hours: 4

Assessment

Group Research Project (45%)

Consists of 2 brief presentations and 1 comprehensive, final presentation. Involves the application of the methods learned to student-led research projects. Students will be asked to apply the roadmap of statistical learning to address a particular question of interest from their own research or that of their fellow students. Project guidelines with details on the required deliverables are posted on bCourses in Files > Project Guidelines.

Present Course Material (20%)

In pairs, students will present a piece of course material (a published paper or vdL&R chapter) and lead a discussion on it for 45 minutes during lecture. Presenters should read the content to be presented and prepare slides in advance of the class meeting. During the presentation, they should briefly summarize the main ideas and key results. Presenters should also prepare a few questions or other discussion points (such as concerns or possible extensions) to encourage discussion and critical thought by other class members. Students will have freedom to decide among themselves which dates they wish to present. Sign-up sheet is posted on bCourses homepage.

Assignments (20%)

Encompass both in-class and out-of-class work compromising theory and data analysis using R. Assignments and assignment due dates will be posted on bCourses.

Blog Question (5%)

In a group of 1 to 3, students will ask a question to the blog by sending an email to vanderlaan.blog@berkeley.edu. The question can be related to (but is not limited to) projects, research, or course content – the criteria is broad, just focus on a statistical question that is not easily “Google-able”.

Attendance (5%)

Student attendance will be recorded at all meetings and will be posted on bCourses.

Course Feedback (5%)

Involves completing an anonymous mid-semester survey and the final course evaluation.

Learning Outcomes

At the end of this course students will 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 to formulate an estimation problem in a survival setting.

Requirements and Materials

There will be assigned and optional readings from the following textbooks that serve as excellent references, encompassing all of the topics covered in this course and more:

- *Targeted Learning: Causal Inference for Observational and Experimental Data* by Mark van der Laan and Sherri Rose (2011) {abbreviated to vdL&R(2011) in bCourses}
- *Targeted Learning in Data Science: Causal Inference for Complex Longitudinal Studies* by Mark van der Laan and Sherri Rose (2018) {abbreviated to vdL&R(2018) in bCourses}
- *The Statistical Analysis of Failure Time Data, Second Edition* by John Kalbfleisch and Ross Prentice (2002)

Mark van der Laan's blog (https://vanderlaan-lab.org/post/) is encouraged for web-based discussion. At any time, students are invited to submit a question to the blog by sending an email to vanderlaan.blog@berkeley.edu. An assignment will require submission of questions to the blog.

<u>The following will be available on bCourses</u>: links to access lecture videos; video lecture notes; assignments; upcoming due dates, the syllabus, attendance, and grades; a sign-up sheet for course material presentations; a research project sign-up sheet and project guidelines.

There will be an anonymous mid-semester feedback survey provided through SurveyMonkey; the link will be sent to students' berkeley.edu email addresses.

Major Deadlines

Project Presentation I – Proposal …………………………………………….... Tuesday, October 15
Project Presentation II – Progress ………………………………………. Thursday, November 14
Blog Question Due …………………………………………......…...…… Tuesday, November 26
Project Presentation III – Comprehensive …….. Thursday, December 12

Course Policies

Late Assignments

Deducted 10% of the total number of points for every late day.

Accommodations

Please speak with the instructional staff as soon as possible if you require any particular accommodations, and we will work out the necessary arrangements.

Scheduling Conflicts

Notify the instructional staff by the second week of the term about any known or potential conflicts (e.g., religious observances, interviews, conferences).

Collaboration and Independence

All assignments should clearly list collaborators and references. Assignments will not be considered for credit if they are a replicate of another classmate's. With that in mind, you may work together but you should complete answers independently.

Honor Code

"*As a member of the Berkeley community, I act with honesty, integrity, and respect for others.*"

The purpose of the Honor Code is to enhance awareness of the need for the highest possible levels of integrity and respect on campus, both within and outside the academic context. We hope and believe that the code will catalyze a series of ongoing conversations about our principles and practices. Through engagement, we can create a consistent message and ethos in our classrooms, labs, departments, and throughout the academic enterprise, to ensure that the core values of academic integrity and honesty are being embraced by both students and faculty.

Academic Integrity

One of the most important values of an academic community is the balance between the free flow of ideas and the respect for the intellectual property of others. Researchers don't use one another's research without permission; scholars and students always use proper citations in papers; and students may not circulate or post materials (handouts, exams, syllabi – any class materials) from their classes without the written permission of the instructor.

Any test, paper or report submitted by you and that bears your name is presumed to be your own original work that has not previously been submitted for credit in another course unless you obtain prior written approval to do so from your instructor. In all of your assignments, you may use words or ideas written by other individuals in publications, web sites, or other sources, but only with proper attribution. If you are not clear about the expectations for completing an assignment or taking a test or examination, be sure to seek clarification from your instructor or GSI beforehand. Finally, you should keep in mind that as a member of the campus community, you are expected to demonstrate integrity in all of your academic endeavors and will be evaluated on your own merits. The consequences of cheating and academic dishonesty – including a formal discipline file, possible loss of future internship, scholarship, or employment opportunities, and denial of admission to graduate school are simply not worth it.