STA 7179 Survival Analysis
Spring 2022

Instructor Hani Doss—222 Griffin-Floyd; email: doss@stat.ufl.edu (email is primarily for administrative purposes, not for questions regarding the course material; for such questions, talk to me during office hours). Zoom Office Hours: MWF period 6, i.e. 12:50pm–1:40pm (if you want to talk to me, please join zoom before 1:20pm). If this time doesn't work for you then let me know (either in class, or by email, or by phone) and I will arrange a meeting at a different time. In-person office hours are at the same time; however, note that social distancing makes in-person office hours in my office somewhat awkward, and I allow only one student in my office at a time. For short questions, you may talk to me in person right after class in the hall. I will email the class the following information, which you should not give out to anyone who is not in the class: my zoom personal ID, and the username and password for the parts of the course webpage that are password protected.

Course Web Page http://users.stat.ufl.edu/~doss/Courses/sa

Course Description This course gives a theoretical development of statistical methods for analyzing life history data, including censored data and truncated data. Topics covered include the Kaplan-Meier estimator, k-sample tests, proportional hazards regression, and the asymptotic theory associated with all these. Throughout, the counting process approach to survival analysis is used.

Prerequisites Prerequisites—in substance, as opposed to by catalog number—are:

- a course in regression at the graduate level;
- a one-year sequence in theoretical statistics at the graduate level;
- a semester of probability at the graduate (Ph.D.) level.

The material in STA 6207 (Regression Analysis), STA 6326-7 (Theoretical Statistics I and II) and STA 6466 (Probability Theory I) are adequate. Regarding the probability prerequisite, it is possible—but difficult—to take this course if you have not had a semester of Ph.D.-level probability but have taken a course in large-sample theory. We will also need some material that is normally covered in the second semester of the Ph.D.-level probability sequence (STA 6467), namely a set of basic facts regarding discrete time martingales, and the central limit theorem (in Lindeberg-Feller form); however, this material will be thoroughly reviewed.

Orientation of the Course This is a Ph.D. elective in the Department of Statistics. A substantial theoretical component is involved. This is not a course on applied survival analysis. If what you are looking for is a course on applied survival analysis you should not take this course.

Texts We will not use any textbooks. We will use the statistical computing language R (which can be downloaded for free from http://www.r-project.org), and a student who is not familiar with it is strongly advised to become so as soon as possible.

Grading Your course grade will be based on the four components below, with the stated weights.

Exam 1:	Friday February 11, 8:20pm.	25%
Exam 2:	Friday March 18, 8:20pm. Covers the material after Exam 1.	25%
Exam 3:	Thursday April 28, 3:00pm–5:00pm. Covers the material after Exam 2.	25%
HW:	There will be about 6 or 7 homeworks assigned during the semester.	25%

Some of the homework assignments will be of a theoretical nature, and some will involve computer implementation of the methods we discuss on specific data sets. The solutions to the homework assignments must be entirely your own (this applies also to R code).

Overview The counting process approach to survival analysis enables us to handle in a unified way a number of different data structures such as random censoring, random truncation, and in general life history data (which arises when individuals are monitored through a series of illness, treatment, wellness, and relapse events). This approach, which will be used in the course, is based on the theory of continuous-time martingales and stochastic integrals, together with the associated martingale central limit theorems. Familiarity with continuous-time martingales will not be assumed; the necessary material will be developed from scratch. The counting process approach involves a different way of viewing and thinking about the data, and this is reflected in the way the computations are done.

Topics

- Overview of models in survival analysis
- Counting processes, stochastic integrals, and martingales
- The one-sample problem and the Kaplan-Meier estimator
- The martingale central limit theorem, and the asymptotic distribution of the Kaplan-Meier estimator
- Comparison of k populations
- Proportional hazards regression. This will take about half the course, and we will discuss the following:
  - The proportional hazards model
  - Partial likelihood
  - Asymptotic distributions and hypothesis tests regarding the regression parameters
  - Residuals and diagnostics
  - Penalized proportional hazards regression