Twin Cities ASA Fall Research Meeting

SHORT COURSE: “Multi-State Survival Models”

Terry Therneau, Mayo Clinic

"Time until event" is a recurring question in medical research: time until death (survival) and the impact of various covariates on that outcome is the traditional example, but the methods apply to any ongoing process: the time until recurrence of cancer, waiting time on the liver transplant list, the onset of comorbidities with aging, the time until return to work (economics) and industrial reliability are a few examples. All are complicated by the presence of censoring, i.e., at the time of analysis, not all subjects will have yet experienced the event. "Survival analysis" is a common label for a set of methods for dealing with this data. Users of survival analysis methods have historically focused on the simple triad of Kaplan-Meier + log-rank + Cox model, all of which focus on a single outcome. And indeed, this has been a very effective tool set for acute disease processes such as advanced cancer where there is a single dominating endpoint of interest. In the last several years, however, multi-state outcome models have become more important in the author's work. A particular example is the Mayo Clinic study of Aging, where there are at least 3 intertwined processes: the progression from none -> mild -> moderate -> severe accumulation of amyloid in the brain, a similar process for the accumulation of tau tangles, and clinical progression from normal -> mild cognitive impairment -> dementia. The course will discuss methods for analyzing such data, motivated by practical examples. To make the results more useful, the course materials have been created as knitr documents so that all of the code to generate the examples will be available to the users. (Computer code itself will be at most a minor portion of the presentation, other than to point out that it isn't that difficult.) Key questions are when and how these tools can be a useful addition to analysis. Topics will include:

- Multi-state "probability in state" curves (Aalen-Johansen)
- Multi-state models
    - Proportional hazards
    - Fein-Gray
    - Interval censoring and parametric models
- Useful predictions from a model
    - Pr(in state s at time t)
    - E(time in state s) or sojourn time
    - Pr(ever enter state s) or lifetime risk

Background: Terry Therneau has been working as a statistician in medical research for over 30 years. He created and maintains the “surivival” package in R, and with Patricia Grambsch is the author of the text "Modeling survival data: extending the Cox model".

INVITED TALKS

"Going Vague: A Collaboration with a Political Scientist and Undergraduate Statistics Researchers to Investigate Ambiguity in Congressional Candidate Issue Statements"

Paul Roback, St. Olaf College

St. Olaf College political science Professor Chris Chapp wished to study the tendency of Congressional candidates to present their political agendas with varying degrees of ambiguity, which can affect voter knowledge of the candidates' policy preferences. Are there certain candidate and district characteristics that are associated with ambiguous issue statements? Can any steps be taken to encourage clearer statements by candidates? As part of a year-long project in St. Olaf's Center for Interdisciplinary Research, Chapp collaborated with three undergraduate statistics students and me to empirically investigate these questions. In this talk, I will describe the challenging and varied dimensions of this project. First, we had to assemble data from varied sources (the US Census, the Cooperative Congressional Election Study, etc.), including scraping 6300 issue statements from 870 candidate websites in the 2014 congressional election. Second, we used supervised machine learning to train and assign ambiguity ratings to all 6300 text statements based on hand-coded statements from the 2012 election. Third, we built multi-level models to explain a candidate's ambiguity ratings using attributes of both districts and individual candidates; we also used hurdle models to explain if a candidate chose to comment on any issues and, if so, how many they chose to issue statements on. Finally, I will share our findings, some of which match political science theory and some of which do not, and their implications for the US political process.

"Bayesian Penalized Regression"

Galin Jones, University of Minnesota

I will consider ordinary least squares, lasso, bridge, and ridge regression methods under a unified framework. The particular method is determined by the form of the penalty term, which is typically chosen by cross validation. The goal is to introduce a fully Bayesian approach which allows selection of the penalty through posterior inference if desired and discuss how to use a type of model averaging approach to eliminate the nuisance penalty parameters. Sufficient conditions for the posterior to concentrate near the true regression coefficients as the dimension grows with sample size will be discussed. Numerical results show that the proposed model and MCMC method tends to select the optimal penalty and performs well in both variable selection and prediction.

"Challenges in Adverse Events Attribution in Cancer Clinical Trials"

Jennifer Le-Rademacher, PhD

Cancer treatments provide a cure or reduce cancer burden for patients. They also can cause severe side effects. The primary goal of cancer treatment trials is to identify effective treatments with minimal toxicity. Reporting of adverse events and their relatedness (known as attribution) to the investigational agents during clinical trials is mandated by regulatory agencies. Determination of whether an adverse event is related to treatment is a challenging and complex process. In this talk, I will give a brief introduction of adverse events collection and attribution assignment in cancer clinical trials. I will discuss the results of our current work evaluating the accuracy of adverse event attribution.

"Statistical Learning of Successful Smiles"

Nate Helwig, University of Minnesota

Individuals with partial facial paralysis suffer severe psychological and social consequences as a result of the inability to create facial expressions such as smiles. Facial reanimation surgery offers hope for these individuals, but clinicians lack a detailed understanding of which features are most important for reanimating socially effective facial expressions. In this talk, I discuss how computer animated facial models and statistical learning methods can provide insight into how people perceive dynamic facial expressions of emotion. Our results reveal that a successful smile involves an intricate balance of mouth angle, smile extent, and teeth show combined with the correct amount of dynamic symmetry. These findings demonstrate the power of statistical learning methods for scientific discovery, and have practical implications for facial reanimation and rehabilitation.