Thursday, May 23, 2019

Location for all the events: The Intellectual House
Address: 4249 Whitman Court, Seattle, WA 98195
Webpage: http://www.washington.edu/diversity/tribal-relations/intellectual-house/
Map: uw.edu/maps/?int

8:30am - 12:30pm: First Short Course

Instructor: Jon Wakefield, Professor of Statistics and Biostatistics, University of Washington

Webpage: http://faculty.washington.edu/jonno/

Email: jonno@uw.edu

Title:
Bayesian Small Area Estimation using Complex Survey Data: Methods and Applications

Abstract:
Small area estimation (SAE) is an important endeavor in global health, epidemiology, and increasingly, in demography. SAE is often based on data obtained from complex surveys, and one must acknowledge the survey design when statistical analysis is performed so that measures of uncertainty incorporate sampling variability and bias is avoided. Often data in particular areas are sparse (perhaps non-existent) and so spatial smoothing is advantageous to 'borrow strength' from neighboring areas.

We will begin with introductions to complex survey data, SAE, spatial modeling, and Bayesian statistics and then bring these topics together to show how reliable SAE estimation can be performed. The SUMMER R package will be used to illustrate ideas.

2:00pm - 5:00pm: Second Short Course

Instructor: Ben Marwick, Associate Professor of Archaeology, University of Washington

Webpage: https://anthropology.washington.edu/people/ben-marwick

Email: bmarwick@uw.edu

Title:

Writing reproducible research with R and Markdown

Abstract:
Let's say you have some data, R code, and a cool result. Now it's time to communicate this with your collaborators (or supervisor). What do you do? In this workshop, we will to show you how to write beautiful, professional reproducible reports, papers and theses using some of the fantastic, free tools and packages made for R. These tools will help you communicate your research, and hopefully mean that you never copy and paste your R output again.

5:30pm - 8:00pm: Conference Mixer and Poster Session

Judge of the Best Poster Contest: Mike Ward, Professor Emeritus of Political Science, Duke University

Friday, May 24, 2019

Location for all the events: Walker-Ames Room, located on the second floor of Kane Hall, in the northeast corner of UW's Central Plaza ("Red Square").
Address: 4069 Spokane Lane, Seattle, WA 98105
Map: uw.edu/maps/?kne

8:15am - 8:45am

Coffee and refreshments

8:45am - 9:00am

Welcome remarks from the conference organizers.

Chair for all the sessions: Adrian Dobra.

First Session: 9:00am-11:40am

9:00am-9:35am

Speaker: Edo Airoldi, Millard E. Gladfelter Professor of Statistics & Data Science, Director, Data Science Center, Fox School of Business, Temple University

Webpage: https://www.fox.temple.edu/mcm_people/edoardo-airoldi/

Email: eairoldi@gmail.com

Title:

Model-assisted design of experiments on social networks and social media platforms

Abstract:
Classical approaches to causal inference largely rely on the assumption of "lack of interference", according to which the outcome of an individual does not depend on the treatment assigned to others, as well as on many other simplifying assumptions, including the absence of strategic behavior. In many applications, however, such as evaluating the effectiveness of healthcare interventions that leverage social structure, assessing the impact of product innovations and ad campaigns on social media platforms, or experimentation at scale in large IT companies, assuming lack of interference and other simplifying assumptions is untenable. Moreover, the effect of interference itself is often an inferential target of interest, rather than a nuisance. In this talk, we will formalize technical issues that arise in estimating causal effects when interference can be attributed to a network among the units of analysis, within the potential outcomes framework. We will introduce and discuss several strategies for experimental design in this context centered around a judicious use statistical models, which we refer to as "model-assisted" design of experiments. In particular, we wish for certain finite-sample properties of the estimator to hold even if the model catastrophically fails, while we would like to gain efficiency if certain aspects of the model are correct. We will then contrast design-based, model-based and model-assisted approaches to experimental design from a decision theoretic perspective.

9:40am - 10:15am

Speaker: Jerry Reiter, Professor of Statistical Science, Duke University

Webpage: http://www2.stat.duke.edu/~jerry/

Email: jreiter@duke.edu

Title:

How Auxiliary Information Can Help Your Missing Data Problem

Abstract:

Many surveys suffer from unit and item nonresponse. Typical practice accounts for unit nonresponse by inflating respondents' survey weights, and accounts for item nonresponse using some form of imputation. Most methods implicitly treat both sources of nonresponse as missing at random. Sometimes, however, one knows information about the marginal distributions of some of the variables subject to missingness. In this talk, I discuss how such information can be leveraged to handle nonignorable missing data, including allowing different mechanisms for unit and item nonresponse (e.g., nonignorable unit nonresponse and ignorable item nonresponse).

10:20am - 10:55am

Speaker: Kosuke Imai, Professor of Government and of Statistics, Harvard University

Webpage: https://imai.fas.harvard.edu

Email: imai@harvard.edu

Title:
Matching Methods for Causal Inference with Time-Series Cross-Sectional Data.

Abstract:
Matching methods aim to improve the validity of causal inference in observational studies by reducing model dependence and offering intuitive diagnostics. While they have become a part of the standard tool kit for empirical researchers across disciplines, matching methods are rarely used when analyzing time-series cross-section (TSCS) data, which consist of a relatively large number of repeated measurements on the same units. We develop a methodological framework that enables the application of matching methods to TSCS data. In the proposed approach, we first match each treated observation with control observations from other units in the same time period that have an identical treatment history up to the pre-specified number of lags. We use standard matching and weighting methods to further refine this matched set so that the treated observation has the outcome and covariate histories similar to those of its matched control observations. Assessing the quality of matches is done by examining covariate balance. After the refinement, we estimate both short-term and long-term average treatment effects using the difference-in-differences estimator, accounting for a time trend. We also show that the proposed matching estimator can be written as a weighted linear regression estimator with unit and time fixed effects, providing model-based standard errors. We illustrate the proposed methodology by estimating the causal effects of democracy on economic growth, as well as the impact of inter-state war on inheritance tax. The open-source software is available for implementing the proposed matching methods.

11:00am - 11:35am

Speaker: Carter Butts, Professor, Departments of Sociology and Statistics, University of California at Irvine

Webpage: http://www.carterbutts.com

Email: carter.butts@gmail.com

Title:
Advances in Exponential Family Random Graph Modeling: Parameterization, Assessment, and Understanding

Abstract:
The development of statistical models for complex networks have been a central challenge within the social sciences, and an area that has seen revolutionary progress in the last two decades. Central to this progress has been the leveraging of statistical exponential families to capture patterns of dependence among edges within networks, leading to the now widely-used paradigm of exponential family random graph (ERGM) modeling. CSSS members (past and present) and their colleagues have played a pivotal role in many of these developments, and at this 20th anniversary meeting it is only natural to consider not only the immediate fruits of those contributions but also some of the new and ongoing advances that they have enabled. This talk will highlight several of these advances, including developments in improved model parameterization, assessment, and model understanding.

11:40am-12:00pm

Coffee and refreshments

Second Session: 12:00pm-1:00pm

12:00pm-12:25pm

Speaker: Matt Taddy, Amazon VP NA Chief Economist, and Professor of Econometrics and Statistics, The University of Chicago Booth School of Business

Webpage: http://taddylab.com

Email: mataddy@amazon.com

Title:
Social Scientific AI

Abstract:

Modern artificial intelligence combines large numbers of machine learning routines together, each solving basic prediction problems within a larger structured application. For example, an AI system to play a board or video game combines distinct machine learning algorithms for each position on the board or game configuration. These AI systems are only possible because their designers know the rules of the game. In more meaningful applications, involving real world problems and multiple human actors, there are no explicit rule books. Instead, we need to look to social scientists to structure problems in a way that they can be broken into many simple prediction tasks, each solvable through machine learning. We need a combination of quantitative social science and statistics and ML to be able to automate and accelerate solutions to real problems with AI.

12:30pm-12:55pm

Speaker: Amanda Cox, Editor of the New York Times's The Upshot section

Webpage: https://en.wikipedia.org/wiki/Amanda_Cox

Email: coxa@nytimes.com

Title:

Design stuff

Abstract:

Precisely what changes when you are making graphics for yourself, for your favorite journal, or for a much broader audience.

Lunch Break: 1:00pm - 2:00pm

Third Session: 2:00pm-3:30pm

2:00pm-2:25pm

Speaker: Kevin Quinn, Professor, Department of Political Science, University of Michigan

Webpage: https://lsa.umich.edu/polisci/people/faculty/kmq.html

Email: kmq@umich.edu

Title:
An Improved Method for Fitting Item Response Theory Models to Sparse Data

Abstract:
Item response theory (IRT) models are widely used by political scientists to measure latent concepts such as attitudes or ideology. Unlike the canonical psychometric applications, many applications in political science feature extremely sparse data as well as a lack of prior information about the sign of the discrimination parameters. This lack of information can result in a posterior density with multiple modes—even when conventional identification assumptions are employed. In such situations, standard model-fitting strategies based on Gibbs sampling and data augmentation do a poor job of exploring all high probability areas of the parameter space. We develop a solution to this problem based on the idea of parallel tempering. We illustrate the approach with examples drawn from political science.

This work is joint with Daniel Ho (Stanford Law School) and Jerry Qiushi Yu (University of Michigan).

2:30pm-2:55pm

Speaker: Derek Kreager, Liberal Arts Professor of Sociology and Criminology, Director, Justice Center for Research, The Pennsylvania State University

Webpage: http://sociology.la.psu.edu/people/dak27

Email: dak27@psu.edu

Title:

A Mixed Methods Approach to Status in Male and Female Prison Units

Abstract:

This presentation overviews results from parallel projects conducted in male and female prison units. Specifically, I focus on gender variation in prison status systems by comparing inmate narratives and network data collected in a women's prison unit (N=135) to results from previously-collected data in a men's unit (N=205). Following an abductive mixed methods design, I first derive and compare themes for unit status from qualitative narratives collected in both prison units. This process arrived at a stark conclusion: the women, much more than the men, are likely to perceive unit status in both positive (e.g., providing collective goods) and negative (e.g., bullying and manipulative) terms. This observed pattern is then pursued in quantitative network analyses, where status nominations are value coded and analyzed with Exponential Random Graph Models (ERGMs). Preliminary results suggest that the networks of positive status ties are similar across the male and female units, but that the network of negative status ties in the female unit are dominated by women sentenced for serious (violent) offenses, long prison stays, and generally disliked by inmate peers. The sum is a detailed examination of informal organization within gendered prison settings with implications for theories of punishment, prison policies, and prisoner-level behavioral and health outcomes.

3:00pm-3:25pm

Speaker: John Ahlquist, Associate Professor, School of Global Policy and Strategy, University of California, San Diego

Webpage: https://gps.ucsd.edu/faculty-directory/john-ahlquist.html

Email: jahlquist@ucsd.edu

Research talk
Title:
Modeling Asymmetric Relationships from Symmetric Networks

Abstract:
Many bilateral relationships requiring mutual agreement produce observable networks that are symmetric (undirected). However, the unobserved, asymmetric (directed) network is frequently the object of scientific interest. We propose a method that probabilistically reconstructs the latent, asymmetric network from the observed, symmetric graph in a regression-based framework. We apply this model to the bilateral investment treaty network. Our approach successfully recovers the true data generating process in simulation studies, extracts new,

politically relevant information about the network structure inaccessible to alternative approaches, and has superior predictive performance.

3:30pm-4:00pm

Coffee and refreshments

Fourth Session: 4:00pm-5:30pm

4:00pm-4:25pm

Speaker: Leontine Alkema, Associate Professor, Department of Biostatistics and Epidemiology, University of Massachusetts, Amherst.

Webpage: https://people.umass.edu/lalkema/

Email: lalkema@umass.edu

Title:
Sensible Statistics for the Social Sciences: How to estimate a population proportion if data are subject to unknown misclassification error?

Abstract:
Even the simplest of statistical problems, estimating a proportion for a population at a given time, transforms into a statistical challenge when the underlying binary variable is reported with unknown misclassification error. We focus on the task of estimating a population proportion for a set of population-periods when information on misclassification is available for a subset of those population-periods only. We present a Bayesian bivariate hierarchical random walk model to estimate sensitivity and specificity in reporting by population and period. We use the model to estimate misclassification errors in the reporting of maternal mortality in vital registration data in high and middle income countries. We find that, due to cause-of-death misclassification, maternal mortality is underestimated by around one-third on average when maternal mortality is high, with bias decreasing with decreasing mortality. Finally, we introduce a simpler approach to account for misclassification errors in settings where misclassification data are too sparse to generalize to population-periods without data. We apply this approach to estimate contraceptive use in developing countries using data on self-reported usage.
This is joint work with UMass Amherst PhD students Emily Peterson and Chuchu Wei and WHO collaborators Doris Chou, Ann-Beth Moller and Lale Say.

4:30pm-4:55pm

Speaker: Mark Ellis, Professor, Department of Geography, and Director, Northwest Federal Statistical Research Data Center, University of Washington

Webpage: https://geography.washington.edu/people/mark-ellis

Email: ellism@uw.edu

Title:

Immobility and Reapeat Migration in the U.S.: Evidence from IRS Longitudinal Records

Abstract:

This paper utilizes a novel dataset linking individual level IRS and SSA administrative records to examine interstate migration longitudinally. The IRS data we use includes tax filings and information returns. Public IRS migration data only includes filers and their dependents, excluding those who do not file. The addition of information returns mitigates these exclusions, reducing selection issues associated with the public data. We make use of the these linked administrative microdata to calculate the size and assess the socio-demographic characteristics of the immobile population (those who never undertake an interstate migration between 2000 and 2015), the mobile population (those who move across state lines at least once between 2000 and 2015), and the population of repeat movers (those who move more than once during this period) differentiating the latter by type of moves: onward to new states and returns to prior states.   We use the income data to assess the trajectories of household income change by percentile for those age 16 through 31 (from 2000-2015) for those who are immobile, migrate once, or engage in return and onward migration.
This is joint work with Lee Fiorio, UW and Thomas B. Foster, US Census Bureau.

5:00pm-5:25pm

Speaker: Patricia Martinkova, Vice Head, Department of Statistical Modelling, Institute of Computer Science, Czech Academy of Sciences.

Webpage: http://www.cs.cas.cz/martinkova

Email: patricia.martinkova@gmail.com

Title:
Deeper analysis of group disparities in ratings motivated by simulated and real data examples

Abstract:

In this talk we present simulated as well as real data examples motivating deeper analysis of disparities in applicant and student ratings. First, we discuss methods for detection of differential item functioning (DIF). We introduce two cases pointing to the importance of DIF analysis: A simulated example showing that, hypothetically, two groups may have an identical distribution of total scores, yet there may be a DIF and thus potentially unfair item present in the data. Contrary, a real data example is provided, whereby the two groups differ significantly in their overall ability, yet no item bias is detected.

Second, we describe real-data example motivating development of a more flexible model-based estimate of inter-rater reliability (IRR). In this example, IRR calculated on stratified data is not able to detect group differences, while the proposed more flexible model-based approach shows significant difference in IRR when rating internal vs. external applicants.

Finally, we present R package ShinyItemAnalysis providing toy data and interactive features for deeper item-level analysis of ratings and assessments. We argue that interesting datasets presented in an interactive way may activate students, enforce understanding, and motivate research developing more flexible analytic methods.

Conference Dinner: 6:00pm - 8:00pm

Location: The University of Washington Faculty Club
Address: 4020 East Stevens Way, Seattle, WA 98195
Webpage: https://www.uofwclub.org
Map: uw.edu/maps/?fac

After dinner invited speech: Adrian E. Raftery, Boeing International Professor of Statistics and Sociology, University of Washington.

Webpage: http://www.stat.washington.edu/raftery

Email: raftery@uw.edu