SCASA KICK-OFF MEETING MINUTES BY MICHAEL TSIANG, SCASA SECRETARY

22 participants

Start at 10:02am

Talk by ASA President Dr. Kathy Ensor - News from ASA and Community Analytics in Action

- Many different types of statisticians in ASA (academia, industry, etc.)
- Nominate our colleagues for awards
  - https://www.amstat.org/your-career/awards-and-scholarships
- Presidential Initiatives
  - DS and AI
    - Committee on DS and AI
  - CSAB - data science program accreditation
    - ASA National Data Mine ($1.5 million) - NSF Funded
  - Leadership
    - Leadership Institute
    - Community Analytics
- Caucus of Industry Representatives
  - Provide a venue and structure to address issues that arise through the private and public sector setting and are important to industry (bio)statisticians and data scientists
  - https://form.jotform.com/zzlalo/CIR-interest-form
  - Start in January
- CSAB (Computer Science Academic Board)
  - Lead ABET member society for accreditation of degree programs in CS, Cybersecurity, DS, Information Systems, IT, and Software Engineering
  - Now a formal DS program accreditation in the works
  - No formal statistics accreditation process, but something our community could consider
- National Data Mine Network
  - PIs: Mark Ward, Kathy Ensor, Monica Jackson, Donna LaLonde, Talitha Washington
  - Directly fund 300 undergraduate students (100 per year)
  - Activities will intentionally restructure the student learning experience to be teamwork oriented and linked to a broader network across the nation
  - Follow-on Regional Innovation-Engines proposal
- ASA Leadership Institute
  - Inaugural Event - IDEA (Influencing Discovery Exploration and Action) Forum
    - Focus on Climate Change
- 3 ASA Presidents join the US Census Bureau
  - Rob Santos, Sallie Keller, Wendy Martinez
  - ASA was originally created to help support the Census!
- Additional ASA Updates
  - Data Science and Literacy Bill - contact Steve Pierson
- Sloan Foundation funding to assess the health of the federal statistical agencies (non-partisan, developing tools and metrics for measurement)
- Developing relations with medical societies focused on cancer research
- ASA a JEDI society
  - "Professional societies play a critical role in growing a diverse and inclusive professional community. This includes something as basic as the climate at professional meetings to the more subtle challenges of ensuring that policies for publications and dissemination of new science are transparent, fair, and just. I am proud that ASA works hard to meet this societal challenge and serves as a beacon for other professional societies" - K. Ensor
- Share news and highlights
  - https://form.jotform.com/zzlalo/highlights
- Upcoming conferences
  - ICHSP - Jan 9-11
  - CSP - Feb 2-4
  - SDSS - May 23-26
  - JSM - August 5-10 - in-person in Toronto

- Wastewater Surveillance (Drowning in Wastewater)
  - Emerging and viable public health management tool
  - Testing of wastewater surveillance started in May 2020
  - Number one data source in Houston for how to understand covid and other viruses
  - Significant investment - building infrastructure
  - www.nytimes.com/2022/01/19/health/covid-comicron-wastewater-sewage.html
  - Houston has 150 total locations sampled each week
  - Wastewater Treatment plants (WWTPs)
    - 39 WWTPs in Houston
    - Analysis Step 0 - Normalization
      - Measured flow/day to transform to copies/day, then log10
    - Analysis Step 1 - Fit natural cubic spline to each WWTP time series
      - Can incorporate temporal and spatial correlation
    - Analysis Step 2 - Estimate city level trend
      - Can forecast positivity rate
  - Wastewater Epidemiology
    - Scientifically sound and robust
    - Need to understand the complexity
    - Semi-automated
    - Actionable and in real time

Break at 11:06am
Start at 11:16am

Talk by Dr. Jason (ZhongZheng) Niu, USC - Application of Bayesian Kernel Machine Regression for the Mixture Association of Air Pollutants with Biological Aging
- Published article:
- One type of air pollution - BTEX
  - WHO estimates air pollution as one of the most important environmental risk factors, accounting for nine million deaths per year
  - Benzene, toluene, ethylbenzene, xylene (BTEX) are a group of aromatic air pollutants retrieved from fossil fuels
    - A lot of indoor sources of BTEX pollutants
    - Ambient levels of BTEX are closely monitored by the EPA and decreased in the past few years, but continuous bio-monitoring of urinary BTEX metabolites in the US showed increasing trend, possibly from indoor sources
- Telomeres
  - Telomeres are protective DNA tandem repeats at each chromosome's ends, protect the functional DNA from losing during cell replication due to the "end replication problem"
  - Telomere length (TL) shortens in cell replication (mitosis), which could be accelerated by amplified proliferation and increased oxidative stress
  - Critically shortened TL can trigger several senescence-signaling (aging) pathways that cause cellular and organ dysfunctions
  - TL may be an important intermediate marker of health outcomes following air pollution exposure
- Limited studies on association of BTEX with TL, and results were inconsistent
  - BTEX usually present in air simultaneously with their levels interrelated ("mixture")
  - BTEX chemicals may also have non-linear, non-additive joint effects
- Hypothesized that BTEX as a mixture is inversely associated with TL
- Study used a subsample from the NHANES (National Health and Nutrition Examination Survey) data
  - TL measured with qPCR in Dr. Blackburn's lab at UCSF
  - BTEX measured by a passive exposure badge for 72 hours, gas chromatography/mass spectrometry (GC/MS)
  - Covariates were considered based on review of previous literature and analyses of the causal structure using directed acyclic graphs
    - Also controlled smoking exposure status
- Mixture analysis
  - Bayesian kernel machine regression (BKMR) model to flexibly model complex non-linear, non-additive effect of an exposure mixture
  - Markov chain Monte Carlo estimation of various effects
- Results
  - Individual BTEX chemicals models are all insignificant
  - With concomitant BTEX chemicals model, many are significant - but multicollinearity may be an issue
  - BTEX does not have much of an interaction effect with other chemicals
  - Exposure to BTEX mixture may be related to short TL at the low-to-middle range of BTEX in general population of adults who are nonsmokers

Break at 12:01pm
Start at 12:16pm

Talk by Dr. Dylan Friel, UC Riverside - A neutral zone classifier for three classes

- Slides:
  https://docs.google.com/presentation/d/1WMLfHtMM9NXkhcj4u3sytMJddMpaLRKWNCa5swy7iqQ/edit#slide=id.p

- Evaluate classifier by more than just overall misclassification rate
- Introduce the "neutral zone" as a means to lower misclassification rates
  - Delay a final classification, perhaps a misclassification is worse than no classification
  - Establish the neutral zone in the context of two classes when the cost of neutral is known or by controlling false positive rate and false negative rate
- Motivation
  - Comments submitted in student evaluations of teaching
  - Analysis done in Likert scale questions
  - Comments are an important aspect of performance reviews
  - There may be hundreds of reviews, classification can assist reviews by grouping comments by sentiment
- Two classes scenario
  - Positive vs not positive
  - Misclassifications are equally unwanted
  - Often use logistic regression
    - Youden threshold is the u which minimizes FPR(u) + FNR(u)
    - Is there a better threshold to control FPR and FNR?
- Neutral Zone
  - Obtain ROC
  - Select two thresholds, introduce a buffer between decision thresholds
  - Possibly delay an official classification
  - Hard error - a misclassified observation
  - Soft error - an observation that falls within the neutral zone
  - Direct consequence is lowering the FPR and FNR and correct classification rate
- Three classes scenario
  - Positive vs negative vs other
  - Want balance in class-conditional error rates
  - Neutral zone if know the cost of misclassification
  - Look to extend the ideas of the two-class neutral zone to control conditional error rates for use with three classes
  - Predicted class is that with the largest predicted probability, "natural boundaries"
- Three-class Neutral Zone - Single L
  - Buffer of size L in (0, 1) for all decision boundaries
  - If we find L such that P(i | j) <= a_ij then we always have at least one solution
  - No neutral zone if L = 0, all neutral zone if L = 1
  - If multiple solutions, choose L that minimizes neutral rate
- Six L Neutral Zone
  - Choose an L for each pairwise decision boundary
- Fine control over all hard error rates
- Achieve desired hard error rates, possibly balance
- Redeem classifier to become usable
- Limited by natural boundaries

- Classification of Text Comments
  - Word2Vec neural network to predict words, extract word embeddings matrix
  - Obtain a numeric vector for each comment
  - Use the vectors as predictors in multinomial logistic regression
  - Predicted class is the one with the largest probability
  - Validated with 5-fold cross validation

- Application to Student Evaluations at UCSC and UCR
  - Each campus is different
  - Unique model for each campus
  - Able to lower misclassification rates
  - Unable to achieve full balance due to underlying natural boundaries and noisy data
  - Only want to provide assistance to reviewers by sorting and giving a summary
  - Avoid cherry picking of comments

- Summary
  - Evaluate classifiers based on conditional misclassification rates not just overall misclassification rates
  - A neutral zone can be introduced to control conditional hard error rates with the cost of lower correct classification rates and number of observations deemed neutral
  - Usable with any data that provides three predicted probabilities

- Future
  - Pilot at UCR
  - Comparison of campuses
  - Bias analysis

Raffle 1:02pm

Business Meeting 1:09pm

- Congratulations to Shujie Ma and Rebecca Le
  - Shujie Ma - Institute of Mathematical Statistics Fellow in 2022
  - Rebecca Le - ASA Council of Chapters Outstanding SCASA Chapter Service Award in 2022

- SCASA Year 2022 Report
  - Virtual Job Fair - February 25
    - Event leader: Neal Fultz
    - Recruiters: Disney, Urgently, Optum, System
    - 40 students from UCI, USC, UCLA, Cal Poly Pomona, CSULB
  - Regional Statistics Data Visualization Poster Competition - March
    - Event leaders: Rebecca Le, Anna Yu Lee, Joyce Fu
  - Virtual Applied Statistics Workshop - April 23
    - Event leaders: Alex Yu, Olga Korosteleva
- Virtual ISEF - May
  - Event leader: Madeline Bauer
- ASA Traveling Course - October 22
  - Event leader: Rebecca Le, Olga Korosteleva

- SCASA Treasurer’s Report
  - 12/9/2022 balance $21,714.41 = $18,395.13 + $3,319.28 ASW money
  - Poster contest -$1017.15
  - ASA Traveling course -$410.62
  - SCASA Membership dues +$2,102
  - Amazon Smile +$73.64
  - Books for raffles at ASW and Kick-Off were paid using our $1000 stimulus money

- 2023-2024 SCASA Board Election Nominees
  - President - Anna Yu Lee
  - Immediate Past President - Rebecca Le
  - President-Elect - Chong Ho (Alex) Yu (nominee)
  - Treasurer - Olga Korosteleva (incumbent)
  - Secretary - Michael Tsiang (incumbent)
  - VP Professional Affairs - Eric Kawaguchi (nominee)
  - VP Academic Affairs - Yingzhou (Joyce) Fu (incumbent)
  - VP Student Affairs - Jennifer Lee (incumbent)
  - ASA Council of Chapters Representative - Wayne Smith (incumbent)
  - E-Newsletter Editor - Olga Korosteleva (incumbent)
  - Webmaster - Olga Korosteleva (nominee)
  - Need to vote on nominees

- 2023 SCASA Tentative Calendar
  - February - Virtual Job Fair
    - Event leader: Eric Kawaguchi
  - March - Virtual HS Poster Competition
    - Event leaders: Rebecca Le, Anna Yu Lee, Joyce Fu
  - April - Virtual ASW
    - Amazon workshop on cloud computing
    - Event leader: Alex Yu
  - May - Virtual ISEF
    - Sign up for judging for ISEF 2024: https://bit.ly/ISEF2024SaveTheDate
  - October - ASA Traveling Course
    - Event leaders: Anna Yu Lee, Olga Korosteleva
  - December - SCASA Kick-Off
    - Event leader: Anna Yu Lee

Finish at 2:06pm