

SCASA Kick-Off and Business Meeting

Saturday, December 2, 2023

Picture: https://share.icloud.com/photos/0498cekVqvIYTtw6tBQuXVL_A

18 in-person attendees

10:30am – Talk by Jacob Rode: “Meta-analysis as a tool to address complex problems: An example on climate change perceptions and overview using R”

- “Influencing climate change attitudes in the United States: A systematic review and meta-analysis” by Jacob Rode et al., published in the Journal of Environmental Psychology
- Conceptual: Summary effect = weighted average of effect sizes
- Random effects meta-analysis
 - o Assume that there is a distribution of true effects
 - o Restricted maximum likelihood (REML)
 - o Hedges’ g
- OSF link: <https://osf.io/u2tcr>
- metafor package in R used for meta-analysis
- Moderator analysis
- Meta-regression for non-independent effects
 - o robumeta package in R (Hedges et al., 2010)
- People’s support for climate policy is hard to change
- Test pairwise differences between levels of moderator
 - o Multiple testing problem
- Results
 - o Moderators: Intervention direction type of intervention, type of attitude
 - o Additional moderator emerged: negative interventions
 - Misinformation
 - Anti-climate change argument
 - Easier to get people to go against climate change than for it
- Takeaways
 - o Climate change attitudes are resistant to change
 - o Skepticism may come easier than belief and action
 - o Policy action is difficult
- Potential pitfalls
 - o Choosing the right meta-analytic approach
 - o Choosing the right working model if you choose RVE
 - o Difficulty in going from coding protocol to dataset
- Benefits
 - o Meta-analysis can help clarify findings from a field of research with mixed results and debates

11:37am – Break

11:55am – Talk by Chong Ho Yu (Alex): “From Traditional Statistics to Modern Data Science and Machine Learning”

- Most of the statistics have remained unchanged for 30 years
- What is happening in the world (in industry)?
 - o Even if you can't change the world, you still need to change with the world!
- Classical Methods
 - o 1805: The least squares criterion for regression modeling
 - o 1892: Pearson's coefficient
 - o 1900: Chi-square analysis
 - o 1908: t-test
 - o 1930: ANOVA
- Model data science and machine learning
 - o 1980s: Convolutional neural networks
 - o 1984: Classification (partition) tree
 - o 1993: C4.5 algorithm
 - o 1997: Gradient boosted tree
 - o 2001: Random forest
 - o 2005: Elastic net
 - o 2006: Adaptive LASSO
 - o 2014: XGBoost
- Hypothesis testing is also known as confirmatory data analysis (CDA)
 - o In 1989, Kenneth Rothman discouraged over-reliance on p-values in the Journal of Epidemiology, but after he left in 2001, the journal reverted to the p-value tradition
 - o The logic of hypothesis testing: Given that the null is true, what is the chance of observing the data in the long run? $P(D|H)$
 - o What we really want to know: Given the data, what is the best theory to explain the data no matter whether the event can be repeated? $P(H|D)$
- Shortcomings of Hypothesis Testing
 - o Do not address treatment effectiveness
 - Example: Study of aspirin for preventing myocardial infarction
 - $n = 22,000$, $p < 0.00001$, the study stopped early due to “conclusive evidence”, aspirin was highly recommended
 - Effect size 0.77%
 - Further research found even smaller effects, and the recommendation was modified
 - o Easy to reject the null
 - o Lack reproducibility
 - o Count on the theoretical distribution
 - o Point estimate
 - o Arbitrary cutoff
 - o Incapable of performing big data analytics
- Bottom Line
 - o Do not suggest eliminating traditional statistics, but we need different tool sets

- What are the differences?
 - Algorithms in the late 20th/early 21st centuries are more efficient than those created in the late 19th/early 20th centuries
 - Better efficiency and accuracy
- Role of computer science
 - In traditional research methods, statistical procedures are incapable of learning from the data
 - AI algorithms can make an adaptive system, i.e., a model that can learn from the new data to improve accuracy
- Role of Database/Data Warehouse Design
 - Data is not always well-structured, and data analysts usually do not receive training in database design.
 - Data is isolated and noncumulative
- Big Data Analytics
 - High volume
 - High velocity
 - High variety
- Common Ground of data science, data mining, big data analytics, etc.
 - Data-driven, not hypothesis-driven
 - Pattern-seeking, not cut-off thinking
 - Multiple approaches (ensemble methods), not a single one
 - Utilize artificial intelligence, not just human judgment
- Advantages of Big Data
 - Behavioral data collected in naturalistic settings (e.g., Google, Facebook) may be more accurate than experimental or survey data
- Common Ground of Classical vs New Methods
- Causal inference

12:50pm – Lunch

1:38pm – Raffle with Olga

1:50pm – Business Meeting

- Year 2023 Report
- Treasurer's Report
 - 12/1/2023 Balance: \$23,047.45
- 2024 Tentative Calendar
 - February-March: Virtual Job Fair
 - March: Data Vis Poster Competition
 - April: Applied Statistics Workshop
 - May: ISEF
 - October: ASA Traveling Course
 - December: SCASA Kick-Off
- ASA DataFest at UCLA

- April 26-28, 2024, UCLA Ackerman Grand Ballroom
- To sign up to be a mentor: <https://forms.gle/WkFbAcEY1GbiEuPc9>
- To sponsor the event: <https://giving.ucla.edu/datafest2024>
- If you would like to be a judge or have questions, can ask Linda Zanonian, linda@stat.ucla.edu.