

SCASA Kick-Off Event: Saturday, December 11, 2021

Total participants: 19

10:06am – Start (Rebecca Le)

10:07am – Celebrating awards and brief comments from award winners

- SCASA Outstanding Service Award 2020: Daniel Jeske, UCR
- SCASA Outstanding Service Award 2020: Robert Gould, UCLA
- Institute of Mathematical Statistics Fellows: Gareth James, USC (not present)
- Distinguished Achievement Award: Janet Myhre (retired)
- ASA-OCLB Outstanding Service Award 2021: Harold Dyck (retired)
- American Statistical Association Fellows: Hongquan Xu, UCLA

10:16am – Robert Gould’s talk on “Data Science Education: Two Initiatives” (Introduction by Rebecca Le)

- What is PreK-12 data science education?
  - There is a great need for DS education
    - Understanding COVID
    - Ethical and political consequences of “big data” and algorithms that scale up
    - Data literacy, data acumen, data fluency, statistical literacy, statistical thinking
  - What is Data Science?
    - An umbrella term, unifying the different sciences
    - Can we just relabel stats?
  - The State of Statistics Education
    - There is no credential for statistics, so teachers usually have math credentials
    - Stats is not always valued by its teachers or school leadership
    - AP Stats has been successful, but relies on calculators, which some feel misrepresents the discipline
    - Outside of AP, statistics often continues to be “plug and chug”
    - Common Core distorted GAISE representations, so the development of statistical ideas in K-12 is weak
  - Another Vision of DS Ed
    - Data are everywhere and accessible to students
    - Wise (2020) A data scientist builds bridges between questions and messy data to make sense of them
- Project 1: DataFest ([datafest.stat.ucla.edu](http://datafest.stat.ucla.edu) and [amstat.org/datafest](http://amstat.org/datafest))
  - A weekend-long celebration of data!
  - A friendly competition for teams of undergraduates to find insight and meaning in a rich and complex data set

- A community of data scientists, statisticians, computer scientists, and mathematicians who support and welcome students
- Get Involved?
  - DataFest 2022 will be held in person at UCLA
  - April 29 – May 1
  - Mentors: <https://forms.gle/U3rbbEeFUsBJ2izi6>
  - Judges by invitation
  - Data?
  - Sponsor!
- Project 2: Mobilize Introduction to DS (introdatscience.org)
  - A year-long HS course in DS
  - Students learn R via RStudio to analyze and wrangle data
  - Engage in participatory sensing to collect data
  - By the numbers:
    - 48 distribution 135 HS participating
    - 28,500 taught to date
  - Emerged from the Mobilize Project, an NSF-funded partnership (2012) between UCLA and LAUSD led by PI Deborah Estrin
  - Goal: Create STEM modules that implement participatory sensing to enhance STEM education, particularly among students in groups underrepresented in sciences
  - Estrin left in 2014 and Rob took over
  - In 2016, the project created IDS
  - The Data Cycle (around a research topic)
    - Ask questions
    - Consider data
    - Analyze data
    - Interpret data
  - Participatory Sensing
    - A data-collection paradigm developed by Estrin’s lab at UCLA (CENS)
    - Mobile devices used to collect data to address various issues: Nutrition, recycling, stress, water conservation
    - Students collect the data: Numbers, pictures, dates, etc.
    - “Human sensors,” collecting a stream of data based on triggers, and not random sampling
    - Builds community
  - Why R (and RStudio?)
    - Communicates statistical models without math notation and within a context that is more meaningful:
      - `mean(~ numberRecycleBins | location, data = trash_campaign)`
    - Interactive, produces results and error messages
    - Ties variables to data sets
    - Develops computational thinking skills, and so ties to CS

- Three advantages to “coding” (Heinzman, E., 2020)
  - Purposeful: not wasting time “memorizing formulas”
  - Efficient: Helps find answers fast
  - Empowering: Gives them control over problem solving
  - Coding was purposeful compared to other math classes
- IDS’s Role in the Curriculum
  - In CA, IDS is a math course, and must be taught by teachers with math credentials
    - Due to recent changes in UC/CSU admissions policies, IDS counts as one year of math towards college admission
    - A viable college-prep sequence could be: Algebra I, Geometry, Data Science, Computer Science
- The Bridge Problem
  - Imagine a group of students who are computer literate, strong programmers able to problem-solve who have a strong understanding of data, but they haven’t had Algebra II
- Context: The Calculus Pathway
  - Algebra I, Geometry, Algebra II, Pre-Calculus, Calculus
  - Debate is when to start the sequence (race to calculus) and whether this best serves all students
  - Large attrition rate among students who step or fall off the pathway
    - Economic/Racial disparities
- California Math Framework
  - <https://www.cde.ca.gov/ci/ma/cf>
  - Does not mandate what is taught, but offers guidelines for how to teach the state standards
  - Proposed framework has a chapter on data science
  - Encourages districts to create “non-calculus” pathways
  - These pathways almost always include statistics and/or data science
  - Math Wars
    - Stats and DS are the battleground on which the math wars are being fought
    - Because Stats/DS are classified as “math”, students will have to choose a calc path or a stats/DS path
    - Inevitably, the stats/DS path will be portrayed as “non-rigorous” and not for college-bound students
- Key DS Educational Take-homes
  - Universal data literacy is of fundamental importance to our society and culture
  - Data literacy cannot be achieved in a single class, but needs to be integrated throughout the preK-12 curriculum (as suggested by GAISE)
  - All students MUST learn data literacy. Some must also learn calculus
  - The bar for “basic” data literacy is higher than it has ever been

- Students can learn most of what they need to know BEFORE they learn the mathematics that supports it (R code can help with this)
- Several National (and international) preK-12 Data Science Education Organizations/Initiatives
  - 11<sup>th</sup> International Conference on Teaching Statistics (ICOTS11)
    - September 11-16, 2022, Rosario, Argentina
    - <https://icots.info/11/>

11:17am – Break

11:31am – Daniel Jeske’s talk on “Statistics and Data Science at NISS and at ISBIS” and “Sample Size Calculations for Personalized Medicine Applications”

- NISS (National Institute of Statistical Sciences)
  - Delivers high-impact research in science and in public policy by leveraging the rich expertise of its staff with that of its base of affiliated organizations in academia, industry, and government
  - Acts as a neutral, objective expert
  - Acts as an independent research organization
  - Acts as a collaborator
  - Recent research projects
    - Survey and estimation methods for agriculture
    - Cancer biomarker discovery
    - Drug safety
  - NISS started in 1990
  - NISS Affiliate Program
    - Only institutional affiliations, no personal affiliations
    - Brings together statistical, mathematical, and DS professionals from all sectors – academia, industry, government/national lab – to support research, information dissemination, human resource development, and networking
  - Recent NISS Affiliate Events
    - Webinar series: Essential Data Science for Business
    - Webinar series: Statistical and Data Science Research Series
      - Can watch recordings: <https://www.niss.org/>
      - Attendance 300-700
    - Webinar series: NISS-Merck Meet-Ups (biostats/bioinformatics)
    - Webinar series: NISS Affiliate Virtual Career Fairs
    - Webinar series NISS Academic Meet-Ups
    - COVID-Related events
    - Workshop: R and Spark Tools for Data Science Workflows
    - Workshop: Writing Workshop for Early Career Researchers
    - Graduate Student Network
      - First graduate student research conference (June 12-13, 2021, Virtual)

- Students presented oral or poster presentation
  - About 20 students participated
- ISBIS (International Society for Business and Industrial Statistics)
  - Started in 1990 within ISI as a working group in industrial statistics
  - Why join ISBIS?
    - There is a biennial international conference that covers case studies, applications, best practices, research developments
    - An active slate of regional conferences, workshops, and local meetings
    - An international journal, Applied Stochastic Models in Business and Industry, with free online access to ISBIS members
    - A young Statisticians' group (y-BIS) facilitates communications among people starting their careers – jobs network, help desk, social networking
- Sample Size Calculations for Mixture Alternatives in a Control Group vs. Treatment Group Design
  - Joint work with Professor Weixin Yao and PhD student Bradley Lubich
  - Sub-populations of patients within a treatment group
    - Responders vs Non-responders
    - Oncology trials
    - Personalized medicine
  - Sub-populations are not identifiable
  - Motivation: A research wants to use a 5% Wilcoxon rank sum test and have 80% power for detecting a treatment effect in responders that is a shift with magnitude  $\frac{1}{2}$  of the standard deviation of the distribution governing control group responses
    - If there are non-responders in the treatment group, the sample size will be too low to get the correct power
  - Even though Wilcoxon is nonparametric under the null, sample size calculations for power are parametric depending on the alternative
  - The normal alternative is the most conservative for estimating sample sizes
  - Summary
    - Powering studies for comparing two treatments is usually done by assuming that the treatment effect is constant
    - Based on the developing understanding of responders and non-responders, we contend that powering a design under a mixture model for the responses from the treatment group is a “safer” approach
    - The required sample size to detect mixture alternatives can be substantially larger than what is required to detect a pure shift alternative
    - A natural way to report the treatment effect is by an estimate of  $(\theta, \sigma)$

- Ongoing work includes looking into other possible estimators and confidence intervals for (theta, sigma)

12:31pm – Raffle (Olga Korosteleva)

12:45pm – Business Meeting (Rebecca Le)

- Participants: 14
- SCASA Report 2021
  - Virtual Job Fair – Friday, February 26, 2021
    - Event leaders: Neal Fultz, Eric Kawaguchi
    - Recruiters: Google, Uber, LeaseLock
  - Regional Statistics Data Visualization Poster Competition – March 2021
    - Event leaders: Rebecca Le, Anna Yu Lee, Olga Korosteleva
    - First place project: “How do the effects of California wildfires in 2020 compare to previous years’ wildfire?” from Mark Keppel High School
  - Virtual Applied Statistics Workshop – Saturday, April 24, 2021
    - Event leaders: Alex Yu, Olga Korosteleva
    - Workshop topic: “Introduction to Data Science, Machine Learning, and Deep Learning (in R and Python)” by Hui Lin and Ming Li
  - Virtual ISEF – May 2021
    - Participated in judging for the ASA Special Awards in Statistics
  - ASA Traveling Course – Saturday, October 23, 2021
    - Event leaders: Rebecca Le, Olga Korosteleva
    - Topic: “An Introduction to Item Response Theory” by Brian Leventhal from James Madison University
- SCASA Treasurer’s Report (Olga Korosteleva)
  - 12/10/2021 Balance: \$21,709.19 = 18,389.91 + 3,319.28 ASW money
  - DataFest -\$500
  - Poster Contest -\$200
  - Applied Statistics Workshop +\$350 (Amgen +\$800, honorarium -\$600, Eventbrite +\$150)
  - ASA Traveling Course +\$260 (Eventbrite +\$240, ASA -\$500)
  - SCASA Membership dues collected +\$1,445
  - All books for all raffles were paid for with our \$1,000 stimulus money
- New Business
  - Create a scholarship committee (fund: membership revenues, \$1,500 annually)
    - Establish criteria, announce the call for applications, review applications, announce awardees
      - Idea: Combine with best graduate paper and best undergraduate paper contests, give scholarships to winners and make the winners present to SCASA at ASW during lunch or at Kick-Off
  - Create committee on nominations and elections

- Three chapter members, be responsible for developing an election slate for each annual election, for conducting each election, for counting the votes, and for informing the chapter membership about election results
- Find a candidate for ASA Council of Chapters Representative to replace Harold Dyck
  - The main duty is to attend the Council of Chapters meetings at JSM and report back to the chapter on relevant issues. The CoC rep gets emails from ASA and CoC and communicates to chapter officers and members. There is also an annual survey of chapter activities that needs to be submitted. The position is an appointed position, but SCASA has had elections for it. It is a three-year term.
- DataFest Announcement (Robert Gould)
  - Please sign up to be mentors!
    - [datafest.stat.ucla.edu](http://datafest.stat.ucla.edu)
    - [amstat.org/datafest](http://amstat.org/datafest)
  - Can email Rob if interested in judging
- ISEF ASA Special Awards in Statistics (Wayne Smith, Madeline Bauer, Harold Dyck)
  - Summary and issues that arose from last year's awards
  - Constructing a judging rubric
  - ASA Special Awards in Statistics Judges offer special recognition and appreciation for Rebecca Le
    - For outstanding contributions above and beyond the call of duty!
  - See the December issue of Amstat News for the first report of Virtual ISEF 2021 [https://magazine.amstat.org/wp-content/uploads/2021/12/December-2021\\_2.pdf](https://magazine.amstat.org/wp-content/uploads/2021/12/December-2021_2.pdf)
- EnCorps STEM Teachers Program (Leah Rodriguez)
  - Empower STEM professionals to transform public education by teaching in under-resourced schools
- New President-Elect for 2023
  - Anna Yu Lee
  - [annayulee1@gmail.com](mailto:annayulee1@gmail.com)
  - <https://www.fuzzybrained.com/collections/all>
- 2022 SCASA Tentative Calendar
  - February 2022: Virtual Job Fair
    - Event leaders: Neal Fultz, Eric Kawaguchi
  - March 2022: Virtual HS poster competition
    - Event leaders: Rebecca Le, Anna Yu Lee, Joyce Fu
  - April 2022: Virtual Applied Statistics Workshop
    - Event leaders: Alex Yu, Olga Korosteleva
    - Need to decide what workshop to hold

- May 2022: Virtual ISEF
- October 2022: ASA Traveling Course
  - Event leaders: Rebecca Le, Olga Korosteleva
- December 2022: SCASA Kick-Off
  - Event leaders: Anna Yu Lee, Rebecca Le