ASA Spring 2025 Meeting Agenda

09:30 – 09:50	Donuts and coffee
09:50 – 10:00	Welcome / Announcements
10:00 – 10:50	Invited Speaker: **Ryan Elmore** *"From the taproom to the classroom: Recent topics in Sports Statistics"*
10:50 – 11:10	**Jason Bernstein**, Kevin Quinlan, Andrew Gillette, Emily Bogle *"Shared Education in Artificial Intelligence and Machine Learning at Lawrence Livermore National Laboratory"*
11:10 – 11:30	**Gina Reynolds** *"Everyday ggplot2 extension: New approaches to light-weight geom extension"*
11:30 – 11:50	**Sean Hardison**, Erin Fedewa, Leah Zacher, James Thorson, Mike Litzow, Franz Mueter *"Projecting seasonal Bristol Bay red king crab distributions by combining predictions from movement and species distribution models"*
11:50 – 13:00	Lunch – sandwiches and drinks provided
13:00 – 13:20	Chapter Business / Elections / David Young Awards / Maurice Davies Awards
13:20 – 13:40	**Sarah Bird**, Ryan Peterson, Michael Kosnett, Ashley Brooks-Russell, Julia Wrobel *"Making highways and workplaces safer: an interpretable machine learning approach to predicting recent cannabis use and impairment"*
13:40 – 14:00	**Erin Hodgess** *"Spatio-Temporal Kriging on the Raspberry Pi with an External GPU"*
14:00 – 14:20	**Jinal Shah**, Nichole Carlson, Debashis Ghosh, Ryan Peterson *"A Practical Guide to Inference After Model Selection in Linear Regression Using R"*
14:20 – 14:40	**Yao Zheng**, Kayleigh Keller *"Measurement Error Correction in Spatial Bayesian Models with Application to Environmental Health"*
14:40 – 15:00	Break
15:00 – 15:20	**Jennifer Chani** *"Analyzing Complaints to Improve Next Generation Medical Devices"*
15:20 – 15:40	**Ksenia Polson** *"Using Student - Generated Case Studies to Learn Ethics in Data Science"*
15:40 – 16:00	**Matthew Simpson**, Christopher Wikle, Scott Holan *"Bayesian Dasymetric Modeling"*
16:00 – 16:20	**Mark Otto** *"Parallel Universes: Decision Analysis and Data Science"*
16:20 – 16:30	Wrapping up
16:45	Drinks/networking at a nearby restaurant/bar (TBD) For those interested in staying, a group will meet for networking/drinks

Location

CU Anschutz Medical Center (AMC) Campus
Anschutz Health Sciences Building
1890 N Revere Ct, Aurora, CO 80045

There is nearby visitor parking in the Vail and Cheyenne Wells parking lots. This campus map includes parking lots. Additional information on visitor parking can be found here.

Contact

Brianna Hitt, ASA Co-Wy President (2024-2025)
Brianna.hitt@afacademy.af.edu

Abstracts

Ryan Elmore, University of Denver Daniels College of Business

"*From the taproom to the classroom: Recent topics in Sports Statistics"*

This presentation concerns several recent studies that fall under the general umbrella of "sports statistics." In particular, I discuss several problems including (1) timeouts in various sports (2) the Houston Astros' cheating scandal, and (3) a noval NBA player shooting metric. I will provide the motivation for each problem, insight into how to find data for problems of this sort, modeling considerations, and results. Lastly, I will highlight several anecdotes with respect to teaching sports statistics.

Jason Bernstein, Lawrence Livermore National Laboratory
Kevin Quinlan, Lawrence Livermore National Laboratory
Andrew Gillette, Lawrence Livermore National Laboratory
Emily Bogle, Portland State University

"Shared Education in Artificial Intelligence and Machine Learning at Lawrence Livermore National Laboratory"

The Shared Education in Artificial Intelligence and Machine learning (SEAM) program at Lawrence Livermore National Laboratory (LLNL) provides AI/ML training to staff with limited background in data science. Participants in SEAM join cohorts focused on a particular topic area such as computer vision, surrogate modeling, and reinforcement learning, and have three months to cover background material on the topic and complete an applied project. This talk begins with a high-level overview of the context, objectives, and logistics of SEAM. The surrogate modeling learning track is then discussed. Surrogate models are statistical or machine learning models, such as Gaussian processes and neural networks, that are used as computationally inexpensive stand-ins for complex computer models. A Bayesian optimization example

with the publicly available inertial confinement fusion (ICF) data set used in SEAM is provided. This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. LLNL-ABS-870394

Gina Reynolds, United States Military Academy at West Point

"Everyday ggplot2 extension: New approaches to light-weight geom extension"

Creating new layers (geom functions) in ggplot2 extension has been described as an 'involved process'. While new educational material has lowered the conceptual barrier to entry, the amount of scaffolding code required for classical extension can make regularly using extension feel prohibitive. In this talk, I'll discuss strategies and tools to make experimentation with new functionality more light-weight and fun. A few practical applications in statistics pedagogy will be presented.

Sean B. Hardison, College of Fisheries and Ocean Sciences, University of Alaska Fairbanks; Shellfish Assessment Program, Alaska Fisheries Science Center, National Marine Fisheries Service, NOAA
Erin Fedewa, Shellfish Assessment Program, Alaska Fisheries Science Center, National Marine Fisheries Service, NOAA
Leah Zacher, Shellfish Assessment Program, Alaska Fisheries Science Center, National Marine Fisheries Service, NOAA
James Thorson, Habitat and Ecological Processes Research Program, Alaska Fisheries Science Center, National Marine Fisheries Service, NOAA
Mike Litzow, Shellfish Assessment Program, Alaska Fisheries Science Center, National Marine Fisheries Service, NOAA
Franz Mueter, College of Fisheries and Ocean Sciences, University of Alaska Fairbanks

"Projecting seasonal Bristol Bay red king crab distributions by combining predictions from movement and species distribution models"

Understanding the seasonal migratory patterns of Bristol Bay red king crab (BB RKC) is key to developing adaptive management strategies that ensure sustainable harvests. However, we currently lack seasonal surveillance of the BB RKC stock distribution outside the late spring, when scientific bottom trawl surveys operate, and autumn, when RKC fishery captains report spatially-resolved crab harvests. Researchers have addressed this challenge by deploying satellite tags on RKC to observe seasonal migratory patterns outside of these seasons, although our capacity to use this information for understanding seasonal stock distributions has been limited. We developed a model of BB RKC movements from satellite tagging data, and used this model to project the distributions of the BB RKC stock from late spring to autumn for each year between 2005-2023. We compared our projections with historical distributions of fishery catches and found that projections skillfully predicted interannual fishery catch distributions. These findings show that movement-informed projections of BB RKC are a useful fishery independent approach to understanding BB RKC seasonal distributions.

Sarah M. Bird, Department of Biostatistics and Informatics, Colorado School of Public Health
Ryan Peterson, Department of Biostatistics and Informatics, Colorado School of Public Health
Michael Kosnett, Department of Medicine, CU School of Medicine
Ashley Brooks-Russell, Department of Community and Behavioral Health, Colorado School of Public Health
Julia Wrobel, Department of Biostatistics and Bioinformatics, Emory University

"Making highways and workplaces safer: an interpretable machine learning approach to predicting recent cannabis use and impairment"

Introduction Due to a rise in fatal crashes where cannabis use is detected, there is substantial need for objective measures of cannabis-induced impairment. Research has identified promising measures to detect recent cannabis use and impairment, but they are often considered in isolation. Our study aimed to jointly evaluate the predictive potential of various physiological and cognitive measures to determine which are most effective at detecting recent cannabis use and impairment.

Methods Adult participants (N = 125) were recruited for an observational study. Users were asked to consume inhaled cannabis products ad libitum over a 15 minute interval and non-users were used as controls. Participants were assessed pre and post smoking with 1) a tablet-based test battery, 2) handheld pupillometry, and 3) blood cannabinoids at baseline (pre-use), and at 40 and 100 minutes, on average, after the start of inhalation. These measures were used to predict recent cannabis use and impairment via penalized logistic regression. Separate models were fit using data collectable in different contexts (roadside and occupational settings).

Results Models with blood THC (at 40 minutes post use) performed best, predicting recent smoking and impairment with areas under the curve (AUCs) of 0.996 and 0.886, sensitivities of 100% and 83%, and specificities of 99% and 82%, respectively. Models with measures 100-minutes post-use performed worse with AUCs of 0.988 and 0.866, sensitivities of 97.8% and 79.2%, and specificities of 96.7% and 82.3%.

Conclusion Including multiple sources of data can improve prediction of recent cannabis use and impairment, over single sources.

Erin Hodgess, University of Northern Colorado

"Spatio-Temporal Kriging on the Raspberry Pi with an External GPU"

We have successfully attached an external AMD GPU to the Raspberry Pi 5. We are using R along with OpenMP Fortran to run spatio-temporal kriging. We are seeing modest gains on the smaller GPU. This can be an effective tool for researchers with small computing budgets and time constraints.

Jinal Shah, Nichole Carlson, Debashis Ghosh, and Ryan Peterson
Department of Biostatistics and Informatics, Colorado School of Public Health, Aurora, CO

"A Practical Guide to Inference After Model Selection in Linear Regression Using R"

For regression problems in the big data era, inferring which variables are most important among a vast number of those available often leads research teams to perform a model selection procedure, such as stepwise regression or penalized models. However, after model selection produces a "final" model, statistical tests and confidence intervals can be overly optimistic if the selection process is not accounted for. Despite recent theoretical advances in making valid inferences post-model selection, there is a lack of practical guidance and user-friendly software to implement and compare these methods. In this work, we review and elaborate upon some fundamental concepts of post-selection inference and demonstrate their application using our novel R package, 'praticalPSI'. While our package's framework is flexible and extendable to new post-selection methods, we focus on stepwise procedures using AIC/BIC criteria, including forward, backward, and bi-directional stepwise models as well as penalized methods including lasso, ridge, elastic net and MCP. For each procedure, one or more of following post-selection inference methods are available: hybrid ordinary least squares (no adjustment), bootstrap, and selective inference. Finally, for each post-selection inference method, we provide three approaches to handling non-selections: ignoring them, treating them as confident (point-mass) nulls, and treating them as uncertain nulls. We illustrate the implementation of these methods with replicable R code, using a case study from the Heart and Estrogen/Progestin Replacement Study.

Yao Zheng and Kayleigh Keller, Department of Statistics, Colorado State University

"Measurement Error Correction in Spatial Bayesian Models with Application to Environmental Health"

In environmental health research, estimating the effects of pollutant exposure on health outcomes is often complicated by spatial misalignment between locations with environmental data and locations where health outcomes are measured. This misalignment can lead to measurement error when assigning exposure estimates to individuals, particularly in large-scale studies where direct exposure measurements are unavailable. To address this challenge, we propose a two-stage spatial Bayesian modeling framework to correct for such measurement error. The framework integrates spatial prediction with survey-based health modeling and explicitly accounts for uncertainty in the exposure estimates. We use a bootstrap procedure for the measurement error correction. Our main contribution is a modeling approach designed for scenarios where direct pollutant measurements from environmental monitoring stations are unavailable or inaccessible, and exposure information must be derived entirely from spatially predicted grid-level data. This reflects common limitations in real-world studies, where predicted exposure surfaces are used in place of monitor observations due to cost, access, or study design constraints. Simulation studies conducted under a range of spatial structures and sample sizes suggest that the proposed methods can improve estimation accuracy while offering flexibility for application across diverse environmental health contexts. By addressing both measurement error and spatial misalignment, our framework contributes a practical solution for improving inference in studies of environmental exposure and health.

Jennifer Chani, NPD Quality Center of Excellence, Medtronic;
Master of Applied Statistics, Colorado State University

"Analyzing Complaints to Improve Next Generation Medical Devices"

Medtronic is currently developing its next generation electrosurgical generator, used to power minimally invasive surgical devices. Two key challenges of this analysis are pulling meaningful trends from complaint related text, and determining the "right" time period over which to count complaints. The strength of R is helping us provide evidence based solutions to these challenges.

Ksenia Polson, Regis University

"Using Student - Generated Case Studies to Learn Ethics in Data Science"

Using case studies derived from learners' experiences in their lives benefits learning. Case studies were used in ethics courses to foster learners' critical thinking, problem-solving and teamwork skills. Course surveys indicated that most student participants agreed or strongly agreed that student-generate case studies made the course relatable to them. Students, also, agreed that student-generated case studies impacted their learning in a positive way. Overall, this process was carried out in two graduate hybrid ethics in data science courses to engage students in learning and hone skills related to working in groups, leadership, and applying course topics to real world situations. This approach is an appropriate pedagogy for this type of course and student population.

Matt Simpson, SAS Institute
Christ Wikle, University of Missouri
Scott Holan, University of Missouri

"Bayesian Dasymetric Modeling"

Official statistics agencies often publicly release microdata with a rich set of measured variables, but geocoded at a coarse scale of spatial resolution to protect the identity of observed households. On the other hand, many statistical agencies also release tabulated estimates of selected variables at finer spatial scales --- for example the number of households that fall into various race, education, and other demographic categories. This trade-off is ever present in official statistics: data users can have a high degree of spatial resolution, or a rich set of measured variables, but never both. Fortunately data users can use both data sources to downscale the microdata to finer spatial scales. We show that one popular method for downscaling, penalized maximum entropy dasymetric modeling (PMEDM), can be viewed as the posterior mode of a particular model with a particular prior distribution. Additionally, we explore the Bayesian framework for PMEDM, including alternate priors and alternate methods for obtaining PMEDM estimates. We apply our methods to Public-use Microdata Samples and tabulated estimates released by the Census Bureau from the American Community Survey.

Mark C. Otto, US Fish and Wildlife Service (retired) and ASA Volunteer

"Parallel Universes: Decision Analysis and Data Science"

In data science, we identify patterns in data and discover what useful, interesting information we can extract. In decision analysis, we utilize that information to make informed choices. Data science primarily functions as a producer of information, while decision analysis serves as a consumer. By leveraging facilitation and decision analysis skills, we broaden our horizons to engage in project framing and collaborate with our colleagues, domain experts, and organization leaders to support their decision-making processes.

Understanding how our identities and beliefs shape our perspective on a problem is quite challenging. However, we can align conflicting sides on a decision by agreeing on the problem framework, establishing values and objectives, determining the data and model that will guide us, generating alternatives, and considering the trade-offs.

It can be challenging to step outside our own beliefs and biases. It is humbling to listen to other sides openly, but this will make us fuller and more effective statistical and data scientists.