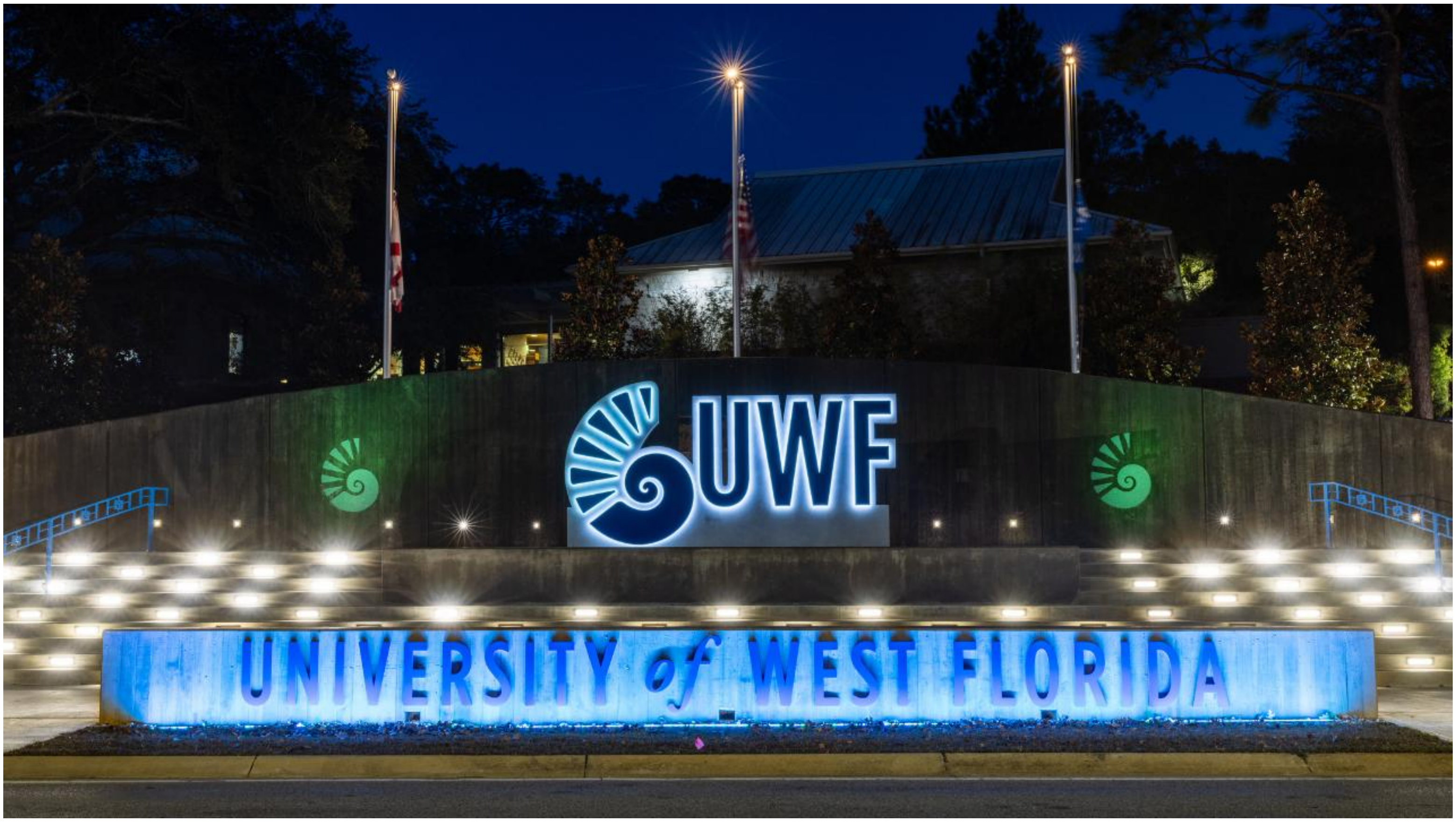
The Southern Regional Council on Statistics
61st Summer Research Conference

June 7–10, 2026
Pensacola, Florida



Welcome to the 61st Summer Research Conference of the Southern Regional Council on Statistics!

The SRC brings together leading researchers, faculty and students in a dynamic, interactive and learning environment.

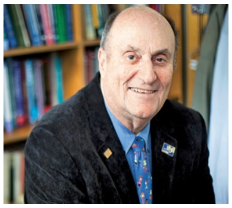
We gratefully recognize these 2026 SRCOS SRC Sponsors.

NSF

ASA
AMERICAN STATISTICAL ASSOCIATION
Promoting the Practice and Profession of Statistics®

NISS
The Statistics Community
Serving the Nation

We also thank Dr. Mike Kutner for his generous contribution.



SRCOS
Southern Regional Council
on Statistics

Conference at a Glance

Southern Regional Council on Statistics
Summer Research Conference 2026 · Pensacola, Florida

Sunday, June 7

5:00 – 7:00 pm	Registration

Monday, June 8

8:15 – 8:30 am	Opening Remarks
8:30 – 10:00 am	Session I: Experimental Design and Discrete Choice Models
10:30 – 12:00 pm	Session II: Small Area Estimation and Distributional Inference
12:30 – 2:00 pm	Poster Sessions
2:00 – 7:30 pm	Free Time & Networking
7:30 – 9:00 pm	Session III: Statistical Education

Tuesday, June 9

8:30 – 10:00 am	Session IV: Machine Learning and AI in Statistics
10:30 – 12:00 pm	Session V: Environmental and Longitudinal Data
12:30 – 5:00 pm	Free Time & Networking
5:00 – 6:00 pm	Reception
6:00 – 8:00 pm	Banquet & Arnold Stromberg Memorial Lecture
8:00 – 8:30 pm	Awards Ceremony

Wednesday, June 10

8:30 – 10:00 am	Session VI: Statistics in Engineering and Public Health
10:30 – 12:00 pm	Session VII: Theoretical Statistics and Causal Inference
12:00 pm	Conference Concludes

SRCOS
Southern Regional Council
on Statistics

Program Schedule

Sunday, June 7, 2026

5:00 – 7:00 pm	Conference Registration

Monday, June 8, 2026

7:30 – 8:15 am	Conference Registration & Continental Breakfast
8:15 – 8:30 am	**Opening Remarks** Katherine Thompson (SRCOS President), University of Kentucky Achraf Cohen (Local Arrangements Chair), University of West Florida
8:30 – 10:00 am	**Session I: Experimental Design and Discrete Choice Models** *Chair: Inkoo Lee, University of Georgia* **Abhyuday Mandal** (Plenary Speaker), University of Georgia *Modeling and Active Learning for Experiments with Quantitative-Sequence Factors* Tharindu Priyan De Alwis, University of West Florida *Dimension Reduction in Matrix Time Series Data*
10:00 – 10:30 am	Coffee Break
10:30 – 12:00 pm	**Session II: Small Area Estimation and Distributional Inference** *Chair: Yin Tang, University of Kentucky* **Gauri Datta**, University of Georgia / U.S. Census Bureau *Small Area Estimation: A Bayesian Approach to Benchmarking and Ranking* Nadeesha Jayaweera, University of Akron *Modeling Dependence in Best-Worst Discrete Choice Data Using a Bayesian INLA Approach*
12:00 – 12:30 pm	Boxed Lunch
12:30 – 2:00 pm	ASA/Kutner Junior and Isolated Faculty Poster Session NSF/Harshbarger Student Poster Session
2:00 – 7:30 pm	Free Time, Networking and Discussion Dinner on your own
7:30 – 9:00 pm	**Session III: Statistical Education** *Chair: D'Arcy Mays, Virginia Commonwealth University* Ryad Ghanam, Virginia Commonwealth University Ed Boone, Virginia Commonwealth University

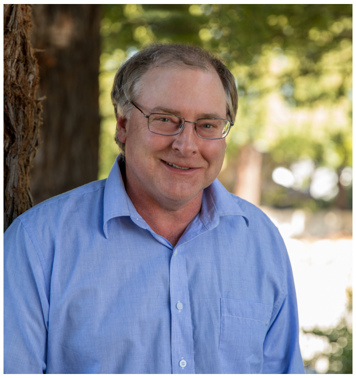
Tuesday, June 9, 2026

7:30 – 8:30 am	Continental Breakfast
8:30 – 10:00 am	**Session IV: Machine Learning and AI in Statistics** *Chair: Garett Ordway, Pensacola Christian College* Qi Zheng, University of Louisville Samuel Kakraba, Tulane University *AI/ML-Enhanced Workflows for Scalable Disease Diagnostics and Quantitative Structure–Activity Relationship (QSAR)–Driven Drug Discovery in Low-Resource Settings* Ruiyan Luo, Georgia State University *SmooNet: Smooth Operator Neural Network and Functional Differential Equation*
10:00 – 10:30 am	Coffee Break
10:30 – 12:00 pm	**Session V: Environmental and Longitudinal Data** *Chair: Ranadeep Daw, University of West Florida* **Robert Lund** (Plenary Speaker), University of California, Santa Cruz *Planetary Snow Cover Changes* Rui (Ray) Xie, University of Central Florida *Adaptive Multi-Agent Feature Selection for Personalized Fall Risk Prevention*
12:00 – 12:30 pm	Boxed Lunch
12:30 – 5:00 pm	Free Time, Networking and Discussion Dinner on your own
5:00 – 6:00 pm	Reception (Cash Bar Available)
6:00 – 8:00 pm	Banquet *Emcee: Mike Kutner, Emory University* **Invited Arnold Stromberg Memorial Lecture** President Montserrat Fuentes, St. Edward's University *Statistics at the Edge of Discovery: AI, Opportunity, and the Future You Can Shape*
8:00 – 8:30 pm	Awards Ceremony and Photos

Wednesday, June 10, 2026

7:30 – 8:30 am	Continental Breakfast
8:30 – 10:00 am	**Session VI: Statistics in Engineering and Public Health** *Chair: Achraf Cohen, University of West Florida* Hakki Sevil, University of West Florida *Entropy-Driven Distributed Modeling and Residual-Based Anomaly Detection for Multi-Agent Systems* Karishma Chhabria Unrue, University of West Florida *Reality of Secondary Data in Public Health: Practical Challenges in Population Science* Mark Carpenter, Auburn University *From Functional to Multivariate Regression: A Unified Operator Perspective with Applications*
10:00 – 10:15 am	Coffee Break (Remember to check out of your room by 11am!)
10:30 – 12:00 pm	**Session VII: Theoretical Statistics and Causal Inference** *Chair: Zhaohu (Jonathan) Fan, Georgia Institute of Technology* Yin Tang, University of Kentucky *Efficient Estimation of the Complier General Causal Effect in Randomized Controlled Trials with One-Sided Noncompliance* **Malay Ghosh** (Plenary Speaker), University of Florida *Bias Corrected Variance Stabilizing Transformation for Small Area Estimation*
12:00 pm	SRCOS 2026 Summer Research Conference concludes *Join us again in 2027 for the next SRC!*

2026 Plenary Speakers



Robert Lund is Professor and Chair, Department of Statistics, University of California, Santa Cruz. He received his applied mathematics undergraduate degree from Auburn University and his PhD in statistics from the University of North Carolina. He is a 2007 elected Fellow of the American Statistical Association and was the 2005–2007 Chief Editor of the *Journal of the American Statistical Association*. He has published over 120 refereed papers and has graduated 26 PhD students. His research expertise lies in probability and statistical climatology, with particular focus on changepoint methods, time series analysis, and the statistical modeling of environmental and climate data. He has received funding from the NSF, NOAA, and NASA for his research on climate trends and extreme weather events. He is also a strong advocate for the next generation of statisticians, regularly serving on editorial boards, NSF review panels, and ASA committees that support early-career researchers.

Planetary Snow Cover Changes

Abstract: This talk develops a statistical model and methods to quantify trends in binary snow cover observations (snow presence/absence, not depths). The methods are used to analyze the Earth's weekly snow cover record, extracted from weekly satellite pictures, during the period 1972–2024. A two state Markov chain with periodic dynamics is developed for the weekly data arising from each land cell studied. Spatial aspects are handled by allowing cell parameters to spatially vary according to a Besag-York-Mollie model. A Bayesian estimation procedure is constructed to estimate model parameters, trends, and their uncertainty margins. Snow cover is seen to be declining in many more cells than it is advancing. While Arctic and European snow cover is found to be rapidly receding, Eastern Canada and Japan are experiencing advancing snow cover. An interesting seasonal pattern also emerges in the estimated trends, shifting the snow season earlier in the winter.

Abhyuday Mandal is a Professor of Statistics at the University of Georgia. He earned his bachelor's and master's degrees from the Indian Statistical Institute, Kolkata, as well as a second master's degree from the University of Michigan, and completed his Ph.D. in 2005 at the Georgia Institute of Technology under the supervision of C. F. Jeff Wu. His research centers on the design of experiments, computer experiments, and optimization techniques, with applications in areas such as drug discovery, small area estimation, agriculture, and engineering. His work has been supported by funding from the National Science Foundation and the National Security Agency, and he has authored numerous influential papers across statistics and applied disciplines. He has mentored many graduate students throughout his career and has been recognized for his contributions to both research and teaching. He also serves as an Associate Editor for several leading journals, including the *Journal of Computational and Graphical Statistics*, *Technometrics*, *Statistics & Probability Letters*, and *Sankhya–B*. In addition, he has been actively involved in the broader statistical community through service on numerous committees and his role in organizing conferences and workshops that bring together researchers from across statistics and related fields.

Modeling and Active Learning for Experiments with Quantitative-Sequence Factors

Abstract: A new type of experiment which targets on finding optimal quantities of a sequence of factors is drawing much attention in medical science, bio-engineering and many other disciplines. Such studies require simultaneous optimization for both quantities and sequence-orders of several components, which is defined as a new type of factors: quantitative-sequence (QS) factors. Due to the large and semi-discrete solution spaces in such experiments, it is non-trivial to efficiently identify the optimal (or near optimal) solutions using only a few experimental trials. To address this challenge, we propose a novel active learning approach, named as QS-learning, to enable effective modeling and efficient optimization for experiments with QS factors. The QS-learning consists of three parts: a novel mapping-based additive Gaussian process (MaGP) model, an efficient global optimization scheme (QS-EGO), and a new class of optimal designs (QS-design) for collecting initial data. Theoretical properties of the proposed method are investigated and techniques on optimization using analytical gradients are developed. The performance of the proposed method is demonstrated via a real drug experiment on lymphoma treatment and several simulation studies.

Joint work with Qian Xiao, Yaping Wang and Xinwei Deng.

Malay Ghosh is a distinguished statistician whose work has had major impact on Bayesian theory and applications, small area estimation, machine learning, and statistical inference. He earned his Ph.D. in Statistics from the University of North Carolina in 1969 after completing both his B.A. and M.A. in Statistics at the University of Calcutta with first-class first distinction and gold medals. Over his career, he has held faculty appointments at the Indian Statistical Institute, Iowa State University, and the University of Florida, where he became Distinguished Professor. He is a Fellow of the Institute of Mathematical Statistics, the American Statistical Association, and the International Society for Bayesian Analysis, and an elected member of the International Statistical Institute. His many honors include the Jerzy Neyman Medal, the Wilks Award, and the 2023 Deming Lecturer Award, and he has authored more than 300 published papers as well as co-authored two books.

Bias Corrected Variance Stabilizing Transformation for Small Area Estimation

Abstract: Small area estimation models are typically based on the normality assumption of response variables. More recently, attention has been drawn to the transformation of the original variables to justify the assumption of normality. Variance stabilizing transformation of observations serves the dual purpose of reaching closer to normality, as well as known variance of the transformed variables in contrast to the assumption of known variances of the original variables, the latter needed to avoid non-identifiability. However, the existing literature on the topic ignores a certain bias introduced in the seemingly correct back transformation. The present paper rectifies this deficiency by introducing asymptotically unbiased empirical Bayes (EB) estimators of small area means. Mean squared errors (MSEs) and estimated MSEs of such estimators are provided. The theoretical results are accompanied with simulations and data analysis. A somewhat surprising phenomenon is a finding which connects one of our results to the natural exponential family quadratic variance function (NEF-QVF) family of distributions introduced by Morris.

Gauri Datta is a Professor in Department of Statistics, University of Georgia and Mathematical Statistician (part-time) at U.S. Census Bureau at the Center for Statistical Research and Methodology (CSRM). He received his Ph.D. at Statistics, University of Florida, 1990. His research interests include Small Area Estimation, Survey Sampling, Bayesian Analysis, Bayes-Frequentist Interface, Likelihood Inference, Multivariate Analysis, Asymptotic, Linear Models, Development of Non-subjective Priors. Dr. Gauri Datta is named ASA Fellow, elected to the Fellowship in the Institute of Mathematical Statistics and the recipient of the Distinguished Alumni Award from University of Florida.

Small Area Estimation: A Bayesian Approach to Benchmarking and Ranking

Abstract: Model-based inference plays an integral role in small area estimation (SAE). The hierarchical Bayesian (HB) modeling provides a logical and easily implementable approach in SAE. This approach facilitates borrowing of information via non-trivial modeling of data from various sources. HB methods are routinely applied to produce point and interval estimation of small area means. Model-based estimates of small area means are often required to be compatible with data at a higher level of aggregation, more accurate national estimates based on a substantially large sample at the national level, either by policy necessity or to produce robust estimation against possible model violation. Many existing benchmarking methods adjust the regular Bayesian estimates of the small area means to comply with the aggregation requirement, but variances of the benchmarked estimates are still estimated under the regular model. Citing deficiencies such as computational complexity for a degenerate posterior and inability to accurately estimate uncertainty of the benchmark estimates of existing Bayesian benchmarking methods, Sugasawa et al. (2024, arxiv:2407.17848v1) argued that by incorporating the benchmark constraints the regular posterior distribution of the small area means can be perturbed to a new posterior distribution that automatically produces benchmarked SAE and their appropriate associated uncertainty. We use this principled approach to make point and interval estimation of small means satisfying the constraints.

While inference for the small means is very crucial in making important policy decisions by national statistical agencies, reliable ranking of the small areas based on their true means is also a useful policy tool. Ranking produced based on the estimated small area means usually ignores uncertainty in the ranking. We propose a Bayesian solution to this problem based on the abovementioned perturbed posterior distribution. Based on an appropriate subset of samples of the means drawn from this modified posterior distribution we develop a joint distribution of the rank vector. We demonstrate the usefulness of our method to estimate median incomes of three demographic groups from 72 PUMAs in Georgia.

Joint work with Abhyuday Mandal from the University of Georgia and Jerry Maples of the US Census Bureau.

2026 Banquet Speaker

Montserrat “Montse” Fuentes is President of St. Edward’s University in Austin, Texas, and Professor of Mathematics. Since beginning her tenure as the university’s 24th president in 2021, and as its first Hispanic president, she has led a period of strategic growth focused on academic innovation, student success, and research development. Under her leadership, St. Edward’s achieved Carnegie Research designation, expanded academic programs including the School of Health Sciences and its first nursing program, strengthened enrollment and retention, and earned the national Seal of Excelencia for advancing Latino student success.

An internationally recognized scholar in statistics and data science, Dr. Fuentes has authored more than 100 scientific publications and secured over $28 million in research funding, including a $3 million NSF ADVANCE award. She is a Fellow of the ASA and a recipient of the U.S. Environmental Protection Agency’s Scientific and Technological Achievement Award.

Before joining St. Edward’s, she served as Executive Vice President and Provost at the University of Iowa, and previously held leadership roles as Dean at Virginia Commonwealth University and Department Head at North Carolina State University. A first-generation college student, Dr. Fuentes earned dual bachelor’s degrees in mathematics and music from the University of Valladolid and a Ph.D. in statistics from the University of Chicago.

Deeply engaged in her community and profession, she serves on numerous civic, healthcare, and higher education boards. Her leadership is grounded in expanding access to transformative education, advancing research excellence, and strengthening the role of universities as engines of opportunity and community impact.

Statistics at the Edge of Discovery: AI, Opportunity, and the Future You Can Shape

Abstract: This is an extraordinary moment to be a student of statistics. Data are transforming nearly every field, artificial intelligence is changing how we learn and work, and society urgently needs people who can turn complexity into knowledge, uncertainty into insight, and evidence into responsible action. For students and early-career statisticians, the opportunities have never been greater. In this banquet talk, Dr. Montserrat “Montse” Fuentes will share a personal and professional reflection on the promise of statistics, drawing from her journey as a first-generation college student, scholar, mentor, department head, dean, provost, president, and researcher in environmental statistics, public health, climate science, spatial modeling, Bayesian methods, and neuroimaging. Her career has shown her that statistics is not only a powerful discipline, but also a passport into many of the most important questions of our time: How does pollution affect health? How do we understand climate risk? How do we study the brain? How do we build AI systems we can trust? How do we make better decisions when the data are incomplete and the stakes are high? The talk will highlight promising research areas, employment pathways, funding opportunities, and the growing role of statisticians in AI and data science. It will also celebrate the special value of SRCOS as a community where students are seen, mentored, encouraged, and invited into the future of the profession. At its heart, this talk is an invitation: to see statistics not just as a major, method, or career path, but as a way to shape the world with rigor, creativity, courage, and purpose.

Abstracts for Oral Presentations

(In order of appearance)

Tharindu Priyan De Alwis

Department of Mathematics and Statistics, University of West Florida
`tdealwis@uwf.edu`

Dimension Reduction in Matrix Time Series Data

Abstract: Matrix-valued data is commonly collected over time in many scientific fields. However, existing methods for handling such data are limited and often suffer from overparameterization. In response, Chen, Xiao, and Yang introduced the matrix autoregressive (MAR) model as an alternative to traditional time series analysis, which relies on vectorization and vector autoregression frameworks. By preserving the original structure of matrices, the MAR model avoids the loss of valuable column and row information. This approach offers a significant reduction in dimensions and enables explicit interpretations of the data. However, when applied to high-dimensional matrix time series, the MAR model faces challenges due to the large size of the coefficient matrices involved. It struggles to differentiate between relevant and irrelevant information, making it inefficient in extracting relevant information from complex data. To address these limitations, we propose envelope-based MAR (EMAR) models that effectively identify and eliminate irrelevant information. Our proposed EMAR approach achieves substantial efficiency gains in estimation and forecasting by reducing parameters and constructing a link between the mean function and covariance structure. This is achieved by using the minimal reducing subspaces of covariance matrices. We establish the asymptotic properties of our proposed estimators and compare their efficiency and accuracy to existing methods through simulation studies under both normality and non-normality conditions. Furthermore, we provide two real-world applications in economics and business to demonstrate the effectiveness of our approach.

Nadeesha Jayaweera

Department of Statistics, University of Akron
`njayaweera@uakron.edu`

Modeling dependence in best-worst discrete choice data using a Bayesian INLA approach

Abstract: We propose a Bayesian framework for analyzing highly dependent best-worst (BW) choice pairs in discrete choice experiments, where strong cross-alternative dependence and temporal structure pose substantial modeling challenges. In our approach, BW outcomes are represented as directed transitions between latent preference states, allowing dependence between best and worst selections to be modeled directly. Transition counts are specified through a Poisson log-linear model with pair-specific random effects to capture heterogeneity and cross-alternative correlation. Bayesian inference is performed using Integrated Nested Laplace Approximation (INLA), providing scalable and computationally efficient estimation without reliance on Markov chain Monte Carlo. The proposed method is evaluated through simulation studies and a quality-of-life case study inspired by

Flynn et al., and is benchmarked against copula-based transition models. Across a range of sample sizes, the framework demonstrates improved model fit and stability, as measured by DIC and WAIC, while yielding robust and interpretable estimates of latent utilities.

Samuel Kakraba

Department of Biostatistics and Data Science, Tulane University
`skakraba@tulane.edu`

AI/ML-Enhanced Workflows for Scalable Disease Diagnostics and Quantitative Structure–Activity Relationship (QSAR)–Driven Drug Discovery in Low-Resource Settings

Abstract: *Background:* Artificial intelligence (AI) and machine learning (ML) offer powerful tools for both disease diagnostics and early-stage drug discovery, yet integrated, scalable workflows that are equitable and suitable for low-resource settings remain limited.

Objective: To demonstrate how rigorously validated, multi-algorithm AI/ML workflows can (1) enhance noninvasive disease diagnostics through an interactive Alzheimer disease prediction platform and (2) accelerate early-stage leukemia drug discovery using quantitative structure–activity relationship–based modeling, with a particular emphasis on scalability, interpretability, and applicability in data-constrained settings.

Methods: We describe multi-algorithm predictive modeling frameworks for disease diagnostics, including an AI-enhanced R Shiny application for Alzheimer disease that integrates diverse supervised learning algorithms, feature selection, and cross-validation within an interactive interface for clinical decision support. We then outline an AI-driven QSAR pipeline for leukemia inhibitor discovery, encompassing molecular descriptor generation, model training and validation across multiple algorithms, external test-set evaluation, and virtual screening for candidate prioritization. Across both use cases, we emphasize design principles for scalability and equity: model calibration, performance assessment across population subgroups, transparency of model outputs, and deployment considerations in data-constrained environments.

Results: In the diagnostic workflow, the AI-enhanced multi-algorithm R Shiny application for Alzheimer disease achieved a test-set accuracy of 91% (95% CI 0.79–0.97) and an area under the receiver operating characteristic curve (AUC) of 0.94 using a neural network classifier trained on handwriting-derived features from 174 participants (89 Alzheimer disease cases and 85 controls). The platform enabled side-by-side comparison of multiple algorithms and identified a compact set of discriminative features while maintaining balanced sensitivity and specificity suitable for clinical decision support. In the drug discovery workflow, the AI-driven quantitative structure–activity relationship pipeline analyzed 35 thiadiazolidinone analogs with 220 molecular descriptors across 17 machine learning algorithms and 12 performance metrics; ensemble methods, particularly LightGBM and Random Forest, yielded superior predictive performance (for example, LightGBM mean squared error $\approx 0.00063 \pm 0.00012$ on held-out data) and robust structure–activity patterns that guided the design of refined leukemia inhibitor candidates with enhanced antileukemic potential.

Conclusions: AI-enhanced, multi-algorithm frameworks such as the Alzheimer disease R Shiny platform and the QSAR pipeline for leukemia inhibitors show that carefully engineered workflows can deliver high diagnostic accuracy and robust compound activity prediction while remaining transparent and deployable in real-world contexts. By reusing common design principles—ensemble modeling, feature stability analysis, rigorous internal and external validation, and user-centered

interfaces—these approaches outline a practical pathway for integrating disease diagnostics and drug discovery into a unified, scalable AI ecosystem that can support more efficient, equitable decision-making across diverse health-care environments.

Ruiyan Luo

Department of Population Health Sciences, Georgia State University
`rluo@gsu.edu`

SmooNet: Smooth Operator Neural Network and Functional Differential Equation

Abstract: Dynamical systems are often modeled by differential equations, where the ordinary differential equations (ODEs) are most used. One major limitation of the ODE model is that it assumes the derivatives of the system only depend on the concurrent values of the state. This can oversimplify the mechanisms of dynamical systems and fail to model the memory effects. To address this limitation, we propose a general Functional Differential Equation (FDE) model which allows the derivative to explicitly depend on both the current value and a historical segment of the system through an unknown operator defined in the space of all continuous functions on an interval. We propose a family of operators, called the Smooth Operator Neural Network (SmooNet), to approximate the operator in FDE. Distinct from the commonly used neural networks with scalar inputs, a unique feature of the SmooNet is that its hidden layer is a continuum with infinite number of nodes, which we call a hidden string, and the complexity of SmooNet is completely determined by the smoothness of the learnable functions. We establish that the SmooNet can universally approximate the operator in FDE, and the solution to the approximate neural FDE can be uniformly and arbitrarily close to the solution of the original FDE. We propose a novel optimization problem based on moving windows to construct the SmooNet, and then make forecasts using the neural FDE.

Rui Xie

College of Nursing, University of Central Florida
`rui.xie@ucf.edu`

Adaptive Multi-Agent Feature Selection for Personalized Fall Risk Prevention

Abstract: Falls among older adults represent a major public health challenge driven by complex, time-varying interactions across multiple risk domains. Effective fall risk factor identification requires learning from heterogeneous longitudinal data while accounting for sparse and delayed fall-related outcome events. However, existing approaches are largely static and fail to adaptively model evolving, individualized risk factors across modalities and time. We propose PAFIR, a Personalized and Adaptive Feature selection framework for fall risk Identification and pRevention, which formulates adaptive feature selection as a reinforcement learning problem over longitudinal multimodal health data. PAFIR jointly models structural dependencies among correlated assessment variables and temporal dynamics in wearable-derived physical activity data, and learns adaptive selection policies across repeated study visits using reward signals derived from sparse fall incidence outcomes. We apply PAFIR to data from the Physio fEedback Exercise pRogram (PEER) cluster-randomized trial. Experimental results demonstrate that PAFIR more effectively captures longitudinal and structural patterns of feature relevance than state-of-the-art baselines, and enables dynamic, subject-specific

feature selection. By adapting selected features over time, PAFIR supports more timely and personalized fall prevention strategies.

Hakki Erhan Sevil

Department of Intelligent Systems and Robotics, University of West Florida
`hsevil@uwf.edu`

Entropy-Driven Distributed Modeling and Residual-Based Anomaly Detection for Multi-Agent Systems

Abstract: Recent advances in Unmanned Aerial Vehicles (UAVs) have made it feasible for large numbers of agents to be deployed in applications such as search and rescue, disaster response, and surveillance. In multi-agent UAV systems, many existing approaches are inspired by natural systems, which provide effective mechanisms for managing collective behavior. Especially, distributed system architectures have been increasingly adopted in UAV applications, where global objectives are typically optimized while maintaining relatively simple individual agent behaviors. In this work, inspired from nature, an application of entropy-based distributed behavior and exogenous anomaly detection in multi-agent UAV systems is presented as a step toward improving system robustness. The interaction dynamics among agents are formalized through an entropy-driven objective to allow system-level uncertainty. For anomaly detection, a parity space approach is developed in which residuals are generated to represent the difference between expected agent responses and observed behavior. Following detection, an isolation mechanism is introduced to mitigate the influence of agent with anomaly on collective dynamics; thereby overall system performance can be preserved. This approach provides a key contribution: it enables the prevention of performance degradation in the multi-agent distributed behavior model, particularly given that the system relies on pairwise interactions and entropy-based relationships among agents. The proposed framework highlights the role of residual-based inference in distributed models, and demonstrates how statistical methodologies can enhance robustness in multi-agent systems.

Karishma Chhabria Unrue

Department of Public Health, University of West Florida
`kunrue@uwf.edu`

Reality of Secondary Data in Public Health: Practical Challenges in Population Science

Abstract: Secondary data underpins much of modern population health research, providing the scale needed to examine complex health patterns within real-world constraints. These datasets, however, are not designed to answer specific research questions, requiring careful attention to variable definition, measurement consistency, missingness, and complex sampling structure. Misalignment between available data and constructs of interest, along with common analytic shortcomings can affect the validity of findings. These limitations carry direct implications for public health practice and policy, where decisions informed by such analyses may inadvertently reinforce existing disparities by reflecting gaps in underlying data. At the same time, advances in artificial intelligence and machine learning offer new opportunities to integrate and analyze large datasets, but also risk amplifying underlying biases if applied without rigor. This presentation examines these challenges and emphasizes a transparent approach to secondary data analysis to support more reliable inferences

in population science.

Mark Carpenter

Department of Statistics and Data Science, Auburn University
`carpendm@auburn.edu`

From Functional to Multivariate Regression: A Unified Operator Perspective with Applications

Abstract: We present a unified operator framework for functional and multivariate regression that clarifies the relationship between commonly used modeling approaches. The central idea is that regression can be formulated as an integral operator defined with respect to a measure, under which functional, scalar, and multivariate models arise as special cases. In particular, discretized functional data correspond naturally to regression under discrete measures, providing an interpretation of standard multivariate regression as a special case of functional regression. We illustrate the implications of this perspective through simulation studies that highlight how different estimation approaches behave under varying levels of discretization and signal structure, and we discuss extensions to multivariate outputs, which follow naturally within the same framework. The goal is to provide intuition for the unified formulation and demonstrate how it can inform both methodological development and practical model selection.

Joint work with Nicholas Gaubatz.

Yin Tang

Department of Statistics, University of Kentucky
`yin.tang@uky.edu`

Efficient estimation of the complier general causal effect in randomized controlled trials with one-sided noncompliance

Abstract: A randomized controlled trial (RCT) is widely regarded as the gold standard for assessing the causal effect of a treatment or intervention, assuming perfect implementation. In practice, however, randomization can be compromised for various reasons, such as one-sided noncompliance. In this talk, we address the issue of one-sided noncompliance and propose a general estimand, the complier general causal effect (CGCE), to characterize the causal effect among compliers. We further investigate the conditions under which efficient estimation of the CGCE can be achieved under minimal assumptions. Comprehensive simulation studies and a real data application are conducted to illustrate the proposed methods and to compare them with existing approaches.

Abstracts for ASA/Kutner Faculty Posters

---

1. **PyCytoData: A Framework for Statistical and Bioinformatic Computing in Multimodal Single-Cell Data**
   Kaiwen Wang, Data Science, Davidson College

   Abstract: With the recent popularity of single-cell technologies such as Mass Cytometry (CyTOF) and Single Cell RNA Sequencing (scRNA-seq), multimodal data on matched subjects are becoming increasingly viable and common. While many popular packages and pipelines have been introduced for scRNA-seq, few have focused specifically on CyTOF and its integration with existing and emerging technologies. In this work, we expand upon PyCytoData, an alliance of statistical tools for CyTOF data analysis. As part of the framework, we propose CySCI, a Bayesian regularized Graphical Attention Model based on weakly connected cross-modal features and within-subject cell abundance regularization. CySI enables joint modeling of multimodal data via an integrated approach. Its multitasking nature also enables cell typing, differential abundance analysis, cross-batch normalization, and latent embedding. Built on the foundation of PyCytoData, CySCI becomes a key piece in constructing a comprehensive suite of analysis pipelines for single-cell-like data beyond CyTOF.

2. **Bayesian Exponential Copula Models for Tensor Mixed-Type Responses**
   Inkoo Lee, Epidemiology & Biostatistics, University of Georgia

   Abstract: Complex data structures, such as tensors and mixed-type responses, are commonly found in dental medicine. Dental hygienists measured 3 biomarkers for each subject at 28 teeth and 6 tooth sites over multiple time points. The data exhibit several challenging characteristics: (1) mixed response types (binary, continuous) (2) a 4-way tensor response for each subject, and (3) heavy-tailedness in the continuous tensor responses. To address these complexities, we propose a novel joint modeling framework for mixed-type tensor responses using a latent variable. This framework incorporates exponential factor copula models to capture heavy-tail dependence, asymmetry, and time dependence, adjusting for multiple time points with theoretical justification. To facilitate sparsity in multi-way tensor response regression, we introduce a tensor-structured horseshoe prior, which enables the identification of a sparse subset of relevant regression coefficients. The MCMC algorithm is employed for efficient sampling of parameters in a hierarchical model. We demonstrate the efficacy and advantages of the proposed methods over alternatives through simulations and the analysis of periodontal disease data.

3. **Efficient Estimation of the Complier General Causal Effect in Randomized Controlled Trials with One-Sided Noncompliance**
   Yin Tang, Dr. Bing Zhang Department of Statistics, University of Kentucky

   Abstract: A randomized controlled trial (RCT) is widely regarded as the gold standard for assessing the causal effect of a treatment or intervention, assuming perfect implementation. In practice, however, randomization can be compromised for various reasons, such as one-sided noncompliance. In this work, we address the issue of one-sided noncompliance and propose a general estimand, the complier general causal effect (CGCE), to characterize the causal effect among compliers. We further investigate the conditions under which efficient estimation of the CGCE can be achieved under minimal assumptions. Comprehensive simulation studies and a real data application are conducted to illustrate the proposed methods and to compare them with existing approaches.

4. **A comparison of areal basis function approaches to analyze physical health over the United States**
   Ranadeep Daw, Mathematics and Statistics, University of West Florida

   Abstract: "Spatial modeling for areal data often uses basis function representations to capture underlying spatial structure efficiently. While continuous-domain bases such as radial kernels or splines are widely used, areal data – defined on discrete, contiguous regions – benefit from bases constructed directly from the spatial adjacency or contiguity structure. In this study, we systematically compare several classes of areal- specific basis functions, including Moran eigenvector maps (MEM), and conditional autoregressive (CAR) type precision matrices and their eigenvector representations. We evaluate each approach in terms of their ability to capture large- and small-scale spatial patterns, their interpretability, and their computational efficiency. Using a representative areal dataset, we illustrate how different basis constructions influence spatial smoothing and model inference. Additionally, we apply these methods to assess the potential causal effects of various amenities on health outcomes across the United States."

5. **Advancing statistics, curing patients: Two novel two-sample tests for survival data with long-term survivors**
   Durbadal Ghosh, Biostatistics, St Jude Children's Research Hospital

   Abstract: "Advances in cancer therapeutics have led to a growing number of clinical settings with substantial proportions of patients who do not experience the event of interest within typical follow-up windows, termed as long-term survivors (L-TS). In two-sample survival studies, treatment differences may arise from variation in the proportion of L-TS, differences in event time distributions among non-L-TS patients, or both. Standard methods such as the log rank test perform best under proportional hazards but may lose power when L TS are present. When follow-up is limited, even correctly specified L-TS based parametric tests may perform poorly because the data do not clearly distinguish long-term survivors from subjects who have not yet experienced the event. These challenges motivate the development of tests that remain robust across diverse survival scenarios involving L-TS. We propose two novel two sample tests for survival data. The first is an adaptive two stage procedure that assesses cure model appropriateness and selects the corresponding cure based or non cure based statistic. The second, L-TS Max, aggregates evidence across complementary tests by taking the maximum of their component statistics. We report the large sample properties and assumptions of these tests, and their finite sample performance and comparisons with existing competitors. We also illustrate the application of our proposed methodology using reconstructed data from a neuroblastoma clinical trial. ”

6. **Randomized Smart Subset Selection for High-Dimensional Generalized Linear Models with False Positive Control**
   Zhaoxue Tong, Statistics, Florida State University

   Abstract: We introduce a novel variable selection method for high-dimensional generalized linear models, called Randomized Smart Subset Selection (RS3). RS3 combines randomized block-wise model fitting with importance-weighted stochastic subset sampling to efficiently identify active predictors. Two extensions are proposed: (i) RS3+SCAD, which applies penalized regression with the SCAD penalty to a reduced set of candidate variables, improving computational efficiency while adaptively determining model size; and (ii) RS3+FPCS, a subsampling-based procedure for false positive control based on coefficient sign stability. Simulation studies demonstrate that RS3 and its variants achieve high true positive rates with low false discovery, even under moderate predictor correlation. Applications to gene expression data show that RS3+SCAD consistently identifies sparse, biologically interpretable gene sets, outperforming existing methods in terms of prediction accuracy and stability.

7. **Collapsing Categories for Regression with Mixed Predictors**
   Chaegeun Song, Mathematics, Bryn Mawr College

   Abstract: ”Categorical predictors are omnipresent in everyday regression practice: in fact, most regression data involve some categorical predictors, and this tendency is increasing in modern applications with more complex structures and larger data sizes. However, including too many categories in a regression model would seriously hamper accuracy, as the information in the data is fragmented by the multitude of categories. In this paper, we introduce a systematic method to reduce the complexity of categorical predictors by adaptively collapsing categories in regressions, so as to enhance the performance of regression estimation. Our method is based on the pairwise vector fused LASSO, which automatically fuses the categories that bear a similar regression relation with the response. We develop our method under a wide class of regression models defined by a general loss function, which includes linear models and generalized linear models as special cases. We rigorously established the category collapsing consistency of our method, developed an Inexact Proximal Gradient Descent algorithm to implement it, and proved the feasibility and convergence of our algorithm. Through simulations and an application to Spotify music data, we demonstrate that our method can effectively reduce categorical complexity while improving prediction performance, making it a powerful tool for regression with mixed predictors.”

8. **Modeling Dependence in Best-Worst Discrete Choice Data Using a Bayesian INLA Approach**
   Nadeesha Jayaweera, Statistics, University of Akron

Abstract: We propose a Bayesian framework for analyzing highly dependent best-worst (BW) choice pairs in discrete choice experiments, where strong cross-alternative dependence and temporal structure pose substantial modeling challenges. In our approach, BW outcomes are represented as directed transitions between latent preference states, allowing dependence between best and worst selections to be modeled directly. Transition counts are specified through a Poisson log-linear model with pair-specific random effects to capture heterogeneity and cross-alternative correlation. Bayesian inference is performed using Integrated Nested Laplace Approximation (INLA), providing scalable and computationally efficient estimation without reliance on Markov chain Monte Carlo. The proposed method is evaluated through simulation studies and a quality-of-life case study inspired by Flynn et al., and is benchmarked against copula-based transition models. Across a range of sample sizes, the framework demonstrates improved model fit and stability, as measured by DIC and WAIC, while yielding robust and interpretable estimates of latent utilities.

9. **Sample size determination for survival studies with surrogate longitudinal biomarkers**
Abhijnan Chattopadhyay, Biostatistics and Bioinformatics Core, Houston Methodist Academic Institute

Abstract: "In survival studies with longitudinal biomarkers, standard joint modeling approaches often assume that the underlying biomarker trajectory is directly observed. In practice, however, biomarkers are frequently measured with error or observed through noisy surrogate measurements, which can lead to biased inference and suboptimal study design. In this work, we develop a sample size determination and power calculation framework for survival models in which the biomarker trajectory is treated as a latent process inferred from surrogate data. We formulate a joint modeling framework for the latent trajectory, observed measurements, and survival outcome. Our methodology employs an information-based approach that accounts for the unobserved process, with key quantities estimated via Monte Carlo methods. This framework yields sample size expressions for key parameters, including the longitudinal–survival association, the direct covariate effect, and a composite effect capturing both direct and indirect pathways through the latent trajectory. The proposed approach accommodates irregular measurement schedules and varying numbers of surrogate observations. Simulation studies show that ignoring measurement error can substantially underestimate the required sample size, whereas the proposed method provides accurate design calculations across realistic scenarios. An interactive Shiny application is developed to facilitate practical implementation."

10. **Balanced Biomarker Enrichment Design Improves Power for Detecting Treatment Effect Heterogeneity Under Biomarker Misclassification: A Simulation Study**
Bakeerathan Gunaratnam, Bioinformatics and Biostatistics, University of Louisville

Abstract: "Clinical trials are increasingly using biomarkers to determine whether treatments work differently across patient groups. However, we still do not clearly know which biomarker-based trial designs work best when the biomarker is not measured perfectly. This study compares the ability of three clinical trial designs to detect differences in treatment effects due to biomarker stratification: the Randomize-All Design, the biomarker-stratified Interaction Design, and the Balanced Enriched Design. We used computer simulations to study these trial designs under varying biomarker prevalence (20% and 30%), specificity (60%-95%), and sensitivity (60%- 90%). For each scenario, we set a different range of interaction effect sizes ($\tau 1$) from 0 to 2 and compared the accuracy of estimating the interaction effect and the power to test it across the three design methods. Each clinical trial design included simulated data from 1,000 participants. For the Balanced Enriched Design, 3,000 participants were screened to enroll 500 tested biomarker-positive and 500 tested biomarker-negative participants. We used a multiple linear regression model with an interaction between treatment and biomarker to test whether the interaction (i.e., treatment effects) differs significantly from zero. That is, to test whether the treatment effect differs significantly between biomarker-positive and biomarker-negative groups. Statistical power was calculated based on 5,000 simulated trials. The simulation results indicate that the Balanced Enriched Design consistently had higher statistical power than the other two designs, regardless of biomarker accuracy. At low specificity (60%) and moderate sensitivity, all designs exhibited similarly low power, indicating poor performance in the presence of substantial biomarker misclassification. As specificity increased, the Balanced Enriched Design reached 80% power at smaller effect sizes ($\tau 1 \approx$ 0.8-1.0), while Randomize-All and Interaction Designs required effect sizes ($\tau 1 \approx$ 1.0-1.2) to achieve similar power. Overall, the results show that the Balanced Enriched Design, which balances biomarker positives and negatives, can greatly improve the power and efficiency of detecting treatment-effect heterogeneity, particularly when biomarker misclassification is low. "

11. Functional ADMM for the AI Era: Fast and Scalable Function-on-Scalar Regression

Zhaohu(Jonathan) Fan, Information Technology Management, Scheller College of Business, Georgia Institute of Technology

Abstract: In the AI era, modern statistical learning increasingly requires methods that are not only theoretically sound but also computationally scalable. This presentation introduces a new functional augmented alternating direction method of multipliers (ADMM) framework for fitting high-dimensional function-on-scalar regression models, where the response is functional and the number of predictors can be very large. Existing approaches often rely on restructuring functional data so that scalar group lasso algorithms can be applied, but that strategy can be computationally inefficient and does not fully exploit the functional nature of the data. To address this gap, we develop an ADMM algorithm directly in the functional setting over Hilbert spaces, yielding efficient closed-form updates through a combination of ridge-type estimation and coordinatewise soft-thresholding. We also establish strong convergence results for the proposed algorithm. Simulation studies show substantial computational gains relative to existing group-lasso-based implementations, while maintaining the same statistical performance. Overall, this work highlights how treating functional learning problems as genuinely functional can unlock faster, more scalable computation for next-generation data science and AI-enabled analytics.

12. **Directed Acyclic Graph-Informed Vine Copula Modeling of Multivariate Data**

Heranga Rathnasekara, Mathematics and Statistics, Old Dominion University

Abstract: In this paper, directed acyclic graph (DAG)-informed vine copula framework is proposed to describe and study multivariate causal inferences. The dependance among variables is presented under a computational framework that integrates modern statistical and computational tools. The proposed method combines the structural and causal interpretability of DAG with the flexibility of R-vine copulas to capture nonlinear, asymmetric, and tail-dependent associations among variables. In this approach, the DAG is first used to identify conditional dependency relationships and to support causal interpretation under appropriate assumptions for understand connected variables. The learned structure is then guides the construction of the R-vine copula, reducing network complexity while preserving meaningful relationships in the data. The vine copula component is specified through a flexible pair-copula decomposition, which enables the joint dependence structure to be modeled beyond standard parametric assumptions. The framework is evaluated through simulation studies and applied to diabetes biomarker data. Model performance is assessed using measures such as AUC, accuracy and sensitivity, while the dependence patterns are shared with causal interpretation and improved disease structure.

Abstracts for NSF/Harshbarger Student Posters

---

1. **Neural Network Models for Multiplex Group Testing Data**

Yu Huang, School of Mathematical and Statistical Sciences, Clemson University

Abstract: Group testing, where individuals are tested initially in pools, is widely used to screen a large number of individuals for rare diseases. Triggered by the recent development of multiplex assays, screening programs now frequently involve testing individuals in pools for multiple infections simultaneously. In infectious disease screening programs that make use of multiplex group testing, it is often desirable to relate individual-level covariates (e.g., age, gender, symptoms) to the underlying joint infection statuses. However, this task is uniquely challenging; not only are individual infection statuses unobserved, masked by the effects of imperfect testing and the pooling protocol, but the co-occurring diseases often exhibit underlying biological correlations. While regression methodologies that address these issues have been developed, existing approaches generally lack the ability to automatically detect and account for nonlinear associations and higher-order interactions between the covariates and the multiple infection statuses. To address these limitations, we propose a neural network framework for multiplex group testing data that jointly estimates individual-level probabilities for multiple diseases. This framework automatically detects and accounts for nonlinear relationships, high-order interactions, and disease correlations if they are present. We evaluate the performance of our approach through extensive numerical studies and further demonstrate its practical utility using Chlamydia and Gonorrhea multiplex group testing data collected by the State Hygienic Laboratory at the University of Iowa.

2. **Empirical Likelihood for Fair Classification using HGR Maximum Correlation**

Kayode Okunola, Mathematics and Statistics, Georgia State University

Abstract: The training of machine learning algorithms on biased data has created various ethical and legal concerns, as these algorithms can learn, enhance, or reinforce such biases. Societal concern is increasing regarding the potential for machine learning algorithms to produce bias based on sensitive attributes such as gender, ethnicity, age, and weight. Numerous cases of this sort have been observed recently, and a new area of study has emerged to tackle these challenges utilizing EL, a non-parametric method. This research focuses on the application of empirical likelihood in fair classification. The aim is to ensure that the predictive model is independent of a sensitive attribute using HGR maximum correlation as a surrogate for fairness and develops the confidence region of the maximum correlation vector using empirical likelihood. Uncertainty is taken into account while evaluating fairness in our confidence region-based classification fairness constraints. The suggested confidence region can be used to test fairness and impose a fairness constraint by utilizing the significance level as a tool to balance accuracy and fairness. The simulation results demonstrate that our method precisely achieves the intended type 1 error rate while properly balancing the trade-off between accuracy and fairness. Data analysis was conducted to implement the effectiveness of our method and a good accuracy score was achieved.

3. **Quadratic Functional Variable Transformation of the Exponential-Gamma Distribution for Flexible Lifetime Modelling**
Taiwo Ayeni, Mathematics and Statistics, Georgia State University

Abstract: This paper studies the reliability implications of a quadratic functional variable transformation applied to a normalised exponential-gamma baseline distribution. The resulting Quadratic-Transformed Exponential-Gamma (QTEG) distribution is derived and its statistical properties are investigated. Closed-form expressions are obtained for the density, distribution function, hazard rate, survival function, entropy, order statistics, and Fisher information matrix. The normalising constant of the exponential-gamma construction is established, and the corrected baseline is shown to occupy the q = 1/2 position in the generalised Gamma family, with formal proof of statistical distinctness from the standard Gamma. Maximum likelihood estimation is developed with a semi-closed starting value, and the existence and uniqueness of the estimator are established. A Monte Carlo simulation study across three hazard-rate regimes and four sample sizes confirms convergence and well-calibrated coverage properties. The practical usefulness of the model is illustrated using three benchmark lifetime datasets compared with several competing distributions using standard goodness-of-fit criteria. The results indicate that the QTEG distribution provides a flexible two-parameter model for modelling heavily right-skewed lifetime data.

4. **Penalized Robust Divide–and–Conquer Empirical Likelihood for High–Dimensional Competing–Risks Data**
Maxime Bouadoumou, Statistics, Georgia State University

Abstract: We study high-dimensional competing-risks regression under the Fine–Gray subdistribution hazard model when the sample size is large and data are distributed across multiple computing nodes. To address computational and inferential challenges in this setting, we develop a penalized robust divide–and–conquer empirical likelihood (PR– DAC–EL) approach for scalable estimation and inference. The full dataset is partitioned into subsets, and a penalized Fine–Gray estimator is computed on each subset to perform variable selection when the number of covariates diverges with the sample size. A one-step bias-corrected estimator is then constructed for each subset, and the selected variables are aggregated to form a global active set. Statistical inference is conducted through an empirical likelihood built from standardized subset-level estimating contributions. To improve stability under heterogeneous or contaminated subsets, we introduce a bounded-influence transformation that limits the impact of anomalous subset contributions during aggregation. Under regularity conditions, we establish consistent support recovery of the proposed estimator. Moreover, on the oracle event that the aggregated support matches the true active set, the robust empirical likelihood ratio statistic converges to a Wilks- type $\chi 2$ distribution. This result enables valid confidence regions and hypothesis testing without resampling or explicit variance estimation. Simulation studies demonstrate that PR–DAC–EL achieves accurate variable selection and reliable coverage probabilities in high-dimensional settings, while remaining stable under subset-level contamination. Applications to the SEER prostate cancer registry and a NASA reliability dataset illustrate the practical performance and scalability of the proposed method.
Keywords: empirical likelihood; competing risks; Fine–Gray model; divide–and–conquer;

high-dimensional data; robust inference; penalized estimation

5. Beyond Endpoint Measurements: Modeling Phenotypic Trajectories for Climate-Informed Genomic Prediction

Soumodip Pal, Statistics, Oklahoma State University

Abstract: Biological phenotypes are the product of complex developmental processes driven by the continuous interplay between genetic and environmental regulators, both internal and external. However, conventional associative and predictive studies often overlook this dynamic relationship, treating growth and development as static endpoint measurements. In the context of organismal adaptation and climate resilience, this static view limits our ability to dissect how genotypes respond to, and are shaped by, environmental fluctuations over the growth and development of an individual. In this study, we addressed this limitation using a hierarchical Bayesian framework that models tree-ring data as growth trajectories, capturing nonlinear development through a Gompertz function shaped by genomic variation, local environment, and climate anomalies. Each Gompertz parameter—representing maximum growing capacity (A), growth rate (K), and inflection point (Ψ)— is further decomposed into key climate-resilience parameters, such as site effect, overall additive genetic effects, and climate sensitivity throughout growth period. This hierarchical decomposition separates genetic, permanent, and environmental variance, while accommodating uncertainty through fully Bayesian inference, revealing how climate variability shapes growth potential. We used interior lodgepole pine (Pinus contorta var. latifolia Douglas) trees, sampled throughout the growth period from 1986-2017, from half-sib families across four locations in central Alberta, Canada, to assess genetic and environmental regulators of growth. Our current results revealed a significant trade-off (r = -0.6)between growing capacity (A) and growth rate (K), and in a warmer scenario, selection for reforestation should target genotypes with modestly accelerated early-season growth (A ¿ 0.15), particularly during the first 10-15 years of growth ($\Psi \approx 10$ years, from 1986-1996), and genotypes from colder growing sites are likely to experience a reduction in maximum growing capacity (A). Our posterior predictive simulations reproduced individual trajectory accurately: observed growth of ring width for representative trees fell within the 95% posterior intervals, and residuals showed no systematic bias. We further extend posterior predictive simulations to estimate tree-ring growth parameters under the projected climate conditions by integrating genetic covariance with climate variables. This approach enables probabilistic forecasting of growth trajectories for each genotype in unobserved future environments. These resulting predictions can identify genotypes most likely to sustain optimal growth, or achieve earlier maturity, under the projected climate scenarios. Overall, the outcome of this study provides actionable insights for climate-adaptive forestry strategies, supporting proactive genotype selection for orchard parents, reforestation planning, and long-term forest resilience.

6. Embedded Experiment Design Methodology for Informative Contrasts in Sample Surveys and Annual Survey of Manufactures (ASM) Redesigns

Yang Li, Dr. Bing Zhang Department of Statistics, University of Kentucky

Abstract: The U.S. Census Bureau periodically redesigns the Annual Survey of Manufactures (ASM) to improve data quality and operational efficiency in this large-scale establishment survey. This study aims to evaluate whether the proposed redesign improves survey efficiency compared with the previous design. The experimental designs embedded within the proposed redesigns will be applied with informative contrasts in the sample surveys to the ASM. Weighted estimates of selected key variables, including total capital expenditure and salaries & wages, are used to assess differences in point estimates and precision. In addition, survey-weighted linear regression models are employed to examine differences in outcome measures across design conditions. The findings will inform future redesign efforts and offer guidance for optimizing efficiency in federal statistical programs.

7. Model Selection for Zero-Inflated Count Data with Missingness: Applications to TDP-43 in Dementia Research

Adrianne Caudill, Dr. Bing Zhang Department of Statistics, University of Kentucky

Abstract: Zero inflation and missing data are common challenges in dementia research, particularly when analyzing count outcomes. Despite their prevalence, there is limited literature addressing model selection for count data which has both zero inflation and missingness. In this study, counts of transactive response DNA binding protein of 43 kDa (TDP-43) were obtained from 136 volunteers across three regions of the hippocampus following death. Substantial zero inflation and some missingness were present. Missing data was addressed using multiple imputation via the MICE package in R. Within each brain region and for each imputed dataset, four distributions were evaluated for the zero-inflated (ZI) counts: Poisson, Negative Binomial, Conway–Maxwell Poisson, and Generalized Poisson. Model parameters were selected using forward selection within each imputation. The final optimal subset of predictors was determined based on the frequency with which variable subsets were selected across imputations. This optimal subset was then used as predictors to fit all four ZI models within each brain region. Models were compared using various model diagnostics. In addition, a simulation study was conducted to assess the impact of zero inflation, missing data mechanism, and missingness proportion on model selection performance. Results from the simulation study contextualize the conditions under which model selection procedures reliably identify the model under varying levels of zero inflation and missingness.

8. Optimized Instance Alteration for Explaining and Assessing Robustness of Classifiers

Evgenii Kuriabov, Statistics, The Pennsylvania State University

Abstract: In this work, we propose a unified approach for diagnosing misclassification and assessing the robustness of black-box classifiers. Central to our method is an optimization framework that modifies an instance so that the classifier predicts a specified target label, while ensuring that the modification remains easily explainable. The objective function contains two components: an explainability-aware L_0 (XA-L_0) penalty that promotes sparse and interpretable modifications, and a classifier loss objective that steers the perturbed instance toward the desired output. This integrated optimization formulation is used both to identify the underlying causes of misclassification and to evaluate robustness by determining how an instance can change within a tolerance region before being reassigned to another class. To quantify robustness, we introduce the Tolerance Region Confusion Matrix (TOR-Confusion Matrix), which measures a classifier's susceptibility by modeling the class-to-class transition probabilities induced by tolerance-bounded perturbations. We validate the proposed method on both image and tabular datasets, demonstrating its ability to jointly deliver interpretability and robustness assessment.

9. Evaluation of the Utility of Normality Tests in Statistical Practice

Benedict Kongyir, Statistics, Oklahoma State University

Abstract: Normality testing is widely used for assessing whether a given dataset conforms to the normal distribution. While it has been discussed in the literature that pre-testing normality can be problematic, there has not been research quantifying the effects of normality pre-testing on downstream inference. For instance, the normality tests can lack utility due to their sensitivity to the sample size. There is also a concern of selective inference if a choice about downstream tests is made based on the results of the normality tests. This research aims to study and quantify these effects by developing a comprehensive simulation-based framework enabling informed decisions regarding normality testing and downstream test procedures. The framework is implemented in R providing a user interface with R Shiny. It allows users to assess the normality of their data using several existing methods, as well as information on the impact of normality test procedures on their chosen downstream procedure. In addition, our research also explores robust machine-learning models for classifying the probability distribution generating the data into 'normal' and 'non-normal'. We demonstrate that our proposed method generalizes across various sample sizes and distributions and delivers a favorable performance compared to the traditional normality tests.

10. Kernel Methods for Compositional Data

Kelvin Njuki, Statistics, Oklahoma State University

Abstract: High-dimensional data are increasingly common in fields such as microbiome, genomics, and other omics, which are often inherently compositional (nonnegative carrying relative information with a constant-sum constraint). Because compositional data lies in a simplex space, they typically require log-ratio transformations to enable standard Euclidean-based statistical analysis. A previous study explored methods for testing association based on Reproducing Kernel Hilbert Space (RKHS), employing them to explore associations between a univariate binary/continuous predictor and a compositional response. This study developed the Compositional Data Analysis using Kernels (CODAK) framework, a kernel distance covariance (KDC) framework used to test the association between cell-type compositions and important predictors, e.g., disease status. It is powerful even in small sample sizes and effectively controls type I error. However, its key limitations motivated this dissertation. First, CODAK relies on the Aitchison distance, which is based on log-ratios, hence incompatible with zero values. Replacing zeros with a pseudo-count was found suboptimal, and increasing zero-inflation leads to a significant power loss and inflation of type I error rates. Second, CODAK was

mainly developed for univariate predictors, though modern studies often involve multivariate and even compositional predictors. Third, KDC performance under different data-generating distributions and zero-inflation patterns remains insufficiently characterized. This research aims to: (i) systematically evaluate KDC framework in compositional data with and without zeros; (ii) develop a new kernel framework that robustly handles zero-inflated compositional data without substantial power loss; (iii) extend kernel-based association testing to settings where both outcomes and predictors are compositional; and (iv) propose a follow-up method to identify which components drive the detected associations. Together, these contributions will provide a more robust and flexible toolkit for analyzing high-dimensional compositional data.

11. Development and Evaluation of an AI-Enhanced Multi-Model R Shiny Application for Predictive Analytics in Alzheimer's Disease

Edmund Agyemang, Biostatistics and Data Science, Tulane University

Abstract: AI has demonstrated superior diagnostic accuracy compared to medical practitioners, highlighting its growing importance in healthcare. SMART-Pred (Shiny Multi-Algorithm R Tool for Predictive Modeling) is an innovative AI-based application for Alzheimer's disease (AD) prediction using handwriting analysis. The research utilized Principal Component Analysis (PCA) to reduce the dimensionality of handwriting data, subsequently training and assessing ten distinct AI models. Model performance was assessed using accuracy, sensitivity, specificity, F1-score, and ROC-AUC metrics. The DARWIN dataset, comprising handwriting samples from 174 participants (89 AD patients, 85 healthy controls), was used for validation. The neural network classifier achieved an accuracy of 91% with a 95% CI ranging from 0.79 to 0.97 and an AUC of 92% on the test set after identifying the most significant features for AD prediction. These results surpass current clinical diagnostic tools, which typically achieve around 81% accuracy. SMART-Pred's performance aligns with recent AI advancements in AD prediction, such as the Cambridge scientists' AI tool achieving 82% accuracy in identifying AD progression within three years using cognitive tests and MRI scans. The variables "air_time" and "paper_time" consistently emerged as critical predictors for AD across all ten AI models, highlighting their potential importance in early detection and risk assessment. To augment transparency and interpretability, the study incorporated the principles of explainable AI, specifically using SHapley Additive exPlanations (SHAP) values, a state-of-the-art method to emphasize the features responsible for our model's efficacy. SMART-Pred offers non-invasive, cost-effective, and efficient AD prediction, demonstrating the transformative potential of AI in healthcare. While clinical validation is necessary to confirm the practical applicability of the identified key variables, this study contributes to the growing body of research on AI-assisted AD diagnosis and may lead to improved patient outcomes through early detection and intervention.

12. BAYESIAN CLASSIFIER FOR HIERARCHICALLY CORRELATED FUNCTIONAL DATA WITH APPLICATIONS TO AUTISM DETECTION IN CHILDREN THROUGH FACIAL DYNAMICS

Irina Kondaurova, Bioinformatics and Biostatistics, University of Louisville

Abstract: Functional data classification methods are gaining popularity across a broad range of disciplines that include engineering, medicine, neuroscience and beyond. The advancement of video data collection technologies enables the development of new biomarkers of neurological disorders based on the behavior and intensity of facial dynamics. In this project, we introduce a novel estimation and classification framework for correlated multilevel functional curves derived from facial Action Unit (AU) intensity data across four emotions (Angry, Happy, Sad, and Surprised). As a proof of concept, we collected and processed facial video data from neurologically typical (NT) children and children diagnosed with Autism Spectrum Disorder (ASD). These data were obtained during human–robot interaction (HRI) sessions, and the resulting analysis provides a foundation for developing an ASD screening or diagnostic tool based on video observations. A Bayesian framework with multiple latent factors was employed to estimate parameters in a functional linear mixed-effects model for correlated multilevel functional data. A generalized projection matrix was applied to data curves in a log-likelihood sampling step to reduce dimensionality and speed calculations of classification probabilities. Computational feasibility was ensured by applying a Variational Bayes (VB) inference method via STAN, which is a faster alternative to traditional MCMC-based sampling techniques to obtain parameter estimates. These estimates were then used to classify ASD status though Bayes' rule and to investigate differences between ASD and NT groups. The proposed classifier showed improved performance relative to standard statistical learning methods.

13. **Scan Statistics for Nonhomogeneous Poisson Processes with Bias-Corrected Transformation and Generalized Extreme-Value Calibration for CNV Detection**
Asanka Duwage, Department of Mathematics and Statistics, Mississippi State University

Abstract: Copy number variation may appear as short genomic regions where tumor and normal sequencing reads differ, even when background read intensity is nonconstant across the genome. This work develops scan-statistic methods for detecting such localized changes in nonhomogeneous Poisson processes. Two related approaches are considered. The first is a bias-corrected transformation-based method that uses a bias-corrected kernel estimator, along with a bandwidth selected to account for that correction, to reduce transformation error before applying the scan statistic. The second is a generalized extreme-value calibration approach that accounts for dependence among overlapping windows. The null transformation is estimated from the control sample and applied to the test sample; under the ideal transform, the grid-based sliding-window counts form a stationary dependent sequence. When the unknown null distribution is replaced by a kernel estimator, the resulting feasible scan statistic remains asymptotically equivalent to the ideal one, so the same extreme-value limit can be used for calibration. Estimating the GEV parameters and the extremal index then provides a practical procedure for calculating critical values. Performance is evaluated under linear and exponential intensity models through Type I error and power, and the methods are applied to tumor-normal sequencing data. The results show that both approaches provide practical tools for scan-based detection under a nonhomogeneous structure.

14. **Stochastic EM for Multistate Models of HIV Progression with Interval-Censored**

Longitudinal Data.
Babatunde Aluko, Statistics, University of Kentucky

Abstract: ABSTRACT

The International Epidemiology Databases to Evaluate AIDS (IeDEA) is a global research consortium that provides extensive HIV/AIDS data from sites around the world. In this study, we propose multistate models (MSMs) to characterize HIV progression across clinical stages while addressing data complexities such as interval-censored and clustered event-history data. To reduce computational burden, we develop a Stochastic Expectation-Maximization (Stochastic EM) algorithm for model estimation. We use simulation to evaluate the performance of the proposed methods and apply the approach to Central Africa IeDEA data to assess the impact of the World Health Organization's 2015 Treat-All policy on HIV disease progression.

15. **Comparison of Causal Network Inference Methods on the Relationship Between DNA Methylation and Transcription**
Hettige Fernando, Bioinformatics and Biostatistics, University of Louisville

Abstract: DNA methylation, a universal epigenetic mechanism, is pivotal in regulating transcription and suppressing gene expression in various ways, and its interference is associated with numerous complex diseases. Graphical networks illustrate the statistical dependence among multiple variables and are widely used in biology, such as gene regulatory networks. In this study, we explored and compared two causal network inference applications to analyze the relationship between DNA methylation and transcription. To achieve this, we generalized the MRTrios package to handle different cancer types, aiming to gain insights into the underlying mechanisms of gene regulation. Our analysis involved studying the relationships between transcription and methylation in these cancer types using data from The Cancer Genome Atlas (TCGA) consortium and the Genomics Data Common portal (GDC). The formulated trios consist of the Copy Number Alteration (CNA) of a gene, the expression (E) of the gene, and the methylation (M) of a site located nearby or within the same gene. We then applied MRGN, a novel causal network inference method that considers many confounding variables under the principle of Mendelian randomization. Using the Bayesian Inference, we calculated the posterior probabilities of edges in the inferred models using genetic variants and the identified confounders in trios for one of the cancer types. Comparing the results of the causal networks obtained from the Machine Learning method and Bayesian Learning method, we observe that most of the causal models generated for each trio are similar for both methods with minor differences. Our comparative analysis highlights the strengths and limitations of each causal network inference method in underlying the complex mechanisms of DNA methylation and Transcription. Our findings provide an advancing understand to researchers on selecting appropriate inference methodologies for analyzing regulatory networks.

16. **SuperSurv: A Unified Framework for Machine Learning Ensembles in Survival**

Analysis
Yue Lyu, Biostatistics and Data Science, The University of Texas Health Science Center at Houston

Abstract: This work introduces SuperSurv, a user-friendly R package for building, evaluating, and interpreting ensemble models for right-censored survival data. Although many survival modeling methods are available, existing tools are often model-specific and lack a unified platform for systematically integrating, comparing, and ensembling heterogeneous learners. SuperSurv addresses this gap by providing a unified interface for diverse survival learners, including models that return full survival curves as well as methods that produce only risk scores. All learner outputs are mapped to calibrated survival probability curves on a common evaluation time grid, enabling direct comparison and ensemble construction across heterogeneous model classes. SuperSurv implements stacking of survival models using inverse-probability-of-censoring weighted (IPCW) Brier risk to estimate ensemble weights in the presence of right censoring. The framework integrates hyperparameter tuning, time-dependent benchmarking metrics, and visualization tools for survival model evaluation. In addition, the package provides post-hoc interpretability utilities based on SHAP (SHapley Additive exPlanations) values and supports covariate-adjusted restricted mean survival time (RMST) contrasts through g-computation. By bridging the gap between theoretical rigor and clinical application, SuperSurv offers researchers a comprehensive ecosystem for modern survival analysis. The SuperSurv package is open-source and available on CRAN at https://CRAN.R-project.org/package=SuperSurv. An empirical example using the METABRIC breast cancer dataset demonstrates a complete workflow from model training and benchmarking to explainability and clinically interpretable survival contrasts.

17. **A Bayesian Hierarchical Meta-Analysis Framework for Estimating Mechanistic Model Parameters from Published Summary Statistics**
Shun Rao, Biostatistics, The University of Texas MD Anderson Cancer Center UTHealth Houston Graduate School of Biomedical Sciences

Abstract: Recent advances in Bayesian adaptive clinical trial designs increasingly integrate pharmacokinetic (PK) and pharmacodynamic (PD) data to support early-phase dose finding and optimization. These designs usually use noninformative priors for certain mechanistic PK/PD model parameters, largely due to limited availability of parameter-level information in conventional clinical trial reporting, which typically provides only aggregate PK summaries. Motivated by ponsegromab, a subcutaneously (SC)-administered monoclonal antibody (mAb) currently investigated in phase I and II studies for cancer cachexia, we develop a three-level Bayesian hierarchical meta-analysis model to infer essential PK model parameters for SC-administered mAbs. The model takes as input conventional summary PK statistics, such as area under the concentration-time curve (AUC) or maximum serum concentration (Cmax). The PK parameters of interest include absorption rate constant (Ka), clearance (CL) and bioavailability (F), among others. Under plausible assumptions about the PK model mechanism, the proposed statistical framework enables reconstruction of dynamic PK profiles, including absorption, elimination, half-life, and duration of the drug action. The estimated PK parameters can in turn be used to enable prediction of drug concentration-time profiles

under various dose/schedule settings via conventional pharmacometrics simulations, as well as incorporation in the form of prior distributions to potentially improve existing or novel PK/PD-based early phase dose-finding/optimization trial designs.

18. Combining Changepoint-Copula Modeling with Count Time Series Analysis

Md Iqbal Hossain, Department of Mathematics and Statistics, Old Dominion University

Abstract: Modeling dependence in count time series is complicated by non-stationarity both in marginal distributions and in the underlying dependence structure, as well as by the discrete nature of the data. Traditional correlation measures (e.g., Pearson's ρ) and stationary copula approaches can produce misleading inferences when structural breaks are present. This work introduces a two-stage framework that integrates changepoint detection with copula modeling to address these challenges. The first stage identifies structural breaks in the marginal distributions using a penalized likelihood method suitable for count data. The second stage applies regime-specific bivariate copula models to capture the dependence structure within each homogeneous period, incorporating a probability integral transform to handle discrete observations. The framework is flexible, allowing the copula family and dependence strength to vary across regimes.

The advantages of the approach are demonstrated through an application to annual tropical cyclone (TC) counts across six major ocean basins from 1980 to 2024. Changepoint detection isolates a structural break in the year 2000, dividing the record into two climatically distinct regimes. Marginal storm frequencies shift substantially, most notably a 59% increase in the North Atlantic, and the dependence structure reorganizes: symmetric Gaussian copulas dominate the early period (53% of basin pairs), while the recent period shows a marked increase in asymmetric, tail-dependent copulas (Gumbel rising from 7% to 20%). Moreover, a statistically significant negative dependence emerges between the Southern Pacific and Northern Indian basins (τ = -0.464, p ¡ 0.001) in the post-2000 regime, a teleconnection that is absent in the stationary analysis. By explicitly accounting for non-stationarity, the proposed framework avoids the aggregation artifacts that can arise when structural breaks are ignored, providing a more accurate characterization of evolving multivariate dependencies. The methodology is broadly applicable to any count time series where regime shifts may affect both marginal behavior and cross-series relationships.

19. Mixture of Covariate-Driven Gaussian Graphical Models

Shibo Guo, Statistic, University of Kentucky

Abstract: This project develops a Bayesian mixture graph model for high-dimensional gene expression and spatial transcriptomics data. It treats samples as coming from multiple latent clusters, which can be understood as different cell types or states, and learns a Gaussian graphical model for each cluster to capture conditional dependencies among genes. These network structures are also allowed to vary with covariates such as spatial coordinates. It uses pseudolikelihood to improve scalability in high-dimensional settings, Normal-Gamma shrinkage for sparse regularization of edge parameters, and Gibbs sampling (with some MH steps in parts) to jointly update cluster assignments, mixture weights, and network parameters. Overall, the goal is to simultaneously perform soft clustering and learn cluster-specific gene networks that can change across spatial locations.

20. Statistical meta-analysis to investigate the association of SNP (rs4680) in COMT gene with multiple cancers

Umma Hafsah Himu, Georgia State University, Georgia State University

Abstract: Background: Previous meta-analyses have investigated the association of SNP (rs4680) in COMT gene with different cancers individually as well as collectively. However, in some cases, meta-analysis results with multiple cancers did not support the cancer-specific meta-analyses results. It may be happened due to fewer individual studies and the selection of inappropriate statistical models. To make a consensus decision more accurately, we aimed to determine the association of rs4680 with multiple cancers by taking more individual studies and appropriate statistical models in this meta-analysis.

Methods: We performed a comprehensive statistical meta-analysis to investigate the association of rs4680 in COMT gene with multiple cancers (case: 44,764 and control: 58,544).

Results: The current meta-analysis with each genetic model indicated that the SNP rs4680 is not significantly associated with the overall cancer risk (p-value ¿ 0.05). The meta-analysis under the cancer-specific sub-group suggested that rs4680 is significantly associated with bladder, prostate, and esophageal cancers, whereas the previous meta-analysis of rs4680 with multiple cancers did not mention the risk of prostate cancer. We observed that the risk of esophageal cancer increases, but the risk of prostate cancer decreases due to the presence of A allele of rs4680. The Trial Sequential Analysis (TSA) results also supported the association of rs4680 with the risk of prostate and esophageal cancers.

Conclusion: The findings of this study recommended that the COMT gene may be utilized as the diagnostic and prognostic biomarker for some particular cancers, including prostate and esophageal cancers.

21. Quantifying Neurocognitive Effects of Sleep Stage Architecture in OSA: A Variable Domain Functional Regression Approach

MD MUSHFIQUR RAHMAN, Data Science, University of Mississippi Medical Center

Abstract: The apnea-hypopnea index (AHI), the conventional metric of obstructive sleep apnea (OSA) severity, fails to capture the temporal dynamics of the overnight sleep stage sequence. We applied Variable-Domain Functional Regression (VDFR) to polysomnographic (PSG) data from 1,103 APPLES participants, treating the epoch-by-epoch sleep stage sequence as a continuous, variable-length functional predictor of AHI, compared across three staging granularities: 5-stage (W, N1, N2, N3, REM), 3-stage (W, NREM, REM), and binary (Wake vs. Non-Wake). Functional sleep stage terms were significant across all stages and model structures (all p ¡ 0.001). Lighter stages Wake-fulness, N1, and N2 were positively associated with AHI, while deeper stages N3 and REM were negatively associated, with REM showing the strongest effect. These associations were attenuated in coarser models, highlighting the importance of fine-grained staging. This is, to our knowledge, the first application of VDFR to overnight PSG data in OSA, revealing stage-specific temporal associations with AHI severity invisible to conventional scalar approaches.

22. Comparing the Performance of the Classical and Bayesian Multiple Regression Estimators when $n \leq k$ and error term not well behaved

Naa Adoley Acquaye, Statistics and Research Methods, University of Denver

Abstract: Ordinary least squares (OLS) performs well when classical regression assumptions are satisfied, particularly when the number of observations exceeds the number of parameters and the design matrix is of full rank. However, when $n \leq k$, OLS may become unstable or fail. This study compares the performance of classical OLS and Bayesian multiple regression estimators under these conditions using a real-world dataset. To examine estimator behavior, the analysis reduced the sample size through random sampling and refit both models across settings in which assumptions were satisfied and violated. Results showed that when n¿kand the model was full rank, the two approaches produced similar estimates. In contrast, when $n=k$, OLS produced unreliable estimates with no standard errors, and when n¡k, OLS yielded unstable and limited results due to overfitting. The Bayesian approach remained more robust across these conditions and produced more reliable estimates. These findings suggest that Bayesian method is a more effective option for parameter-rich settings in which classical methods break down.

23. Bayesian Hierarchical Spatiotemporal Prediction of Multi-Pathogen Outbreak: Integrating Mobility and Environmental Exposures.

Hakeem Adekunle, Mathematics and statistics, Georgia State University

Abstract: Introduction: Infectious disease outbreaks remain a persistent global health burden, particularly as populations experience the concurrent circulation of multiple infectious agents that interact across space and time. Forecasting such complex epidemic systems requires models that can capture shared transmission mechanisms, pathogen- specific dynamics, and uncertainty arising from incomplete surveillance.

Methods: In this study, we develop a unified Bayesian hierarchical spatiotemporal framework for predicting multi-pathogen outbreak trajectories while integrating human mobility patterns, environmental exposures, and structured reporting uncertainties. We conducted a comprehensive simulation experiment to evaluate the model's ability to recover known parameters, distinguish pathogen-specific transmission effects, and generate calibrated forecasts under varying levels of reporting noise and spatial heterogeneity. We further applied the method to CDC FluView surveillance weekly data from the United States, spanning January 2017 to December 2025.

24. Longitudinal Changes in Diurnal Physical Activity Trajectories among Older

Adults: Evidence from NHATS Rounds 11–14

Md Ariful Islam Sanim, Statistics, University of South Carolina

Abstract: Purpose Physical activity (PA) is a crucial modifiable risk factor for overall health among older adults. While most studies use PA volume as a summary metric, continuously monitored accelerometer data can capture more detailed daily PA patterns. While the total volume of physical activity generally declines due to aging, the longitudinal evolution of diurnal PA patterns remains understudied. This study investigates longitudinal changes in daily physical activity trajectories among older U.S. adults.

Methods We analyzed continuous 24-hour accelerometer data from the National Health and Aging Trends Study (NHATS), rounds 11–14 (2021–2024), for participants aged 65 years and older. Functional Principal Component Analysis (FPCA) was employed to identify the dominant modes of variation in daily activity. We applied a Projection-Based Functional Invariance Test (PROFIT) to formally test for significant differences in mean activity trajectories over time. Linear mixed-effects models were used to quantify the association of age, sex, and BMI with the longitudinal changes in the principal component scores.

Results The PROFIT test revealed statistically significant changes in diurnal PA patterns across the four rounds. Two primary modes of variation in daily trajectories: the overall magnitude of daily activity (FPC1) and a contrast between morning and evening activity (FPC2) was found to significantly change over the years. Further exploration using pairwise comparisons indicated similarity between Rounds 11 and 12, followed by a significant decline in mid-day activity in Round 13 and a subsequent increase in afternoon and evening activity in Round 14. Mixed-effects modeling found that overall activity magnitude decreased significantly with advanced age, was lower for males, and was inversely associated with BMI.

Conclusions Physical activity trajectories among older adults are dynamic with significant longitudinal evolution in both intensity and timing over a four-year period. With age, the physical activity patterns among older adults may decline due to several factors such as chronic diseases and comorbidity. Our findings highlight the value of high-resolution accelerometry and functional data analysis in uncovering behavioral nuances such as age-related shifts in the daily PA pattern, that summary metrics may overlook. Longer continuous follow up is necessary to understand these patterns properly and make informed public health decisions.

25. A Note on the Gradient-Evaluation Sequence in Accelerated Gradient Methods for Statistical Computing

Yipeng Zhang, School of Statistical and Mathematical Science, Clemson University

Abstract: Accelerated gradient methods are fundamental tools for large-scale optimization problems arising in statistical learning and data analysis. In the classical accelerated gradient framework, two sequences of iterates are generated: one is used for gradient evaluations, while the approximate solution is selected from the other. For unconstrained problems, previous work has shown that the gradient-evaluation sequence can itself serve as the approximate-solution sequence without sacrificing the optimal order of iteration complexity. For projection-based accelerated gradient methods with projection-friendly feasible sets, however, whether the same property holds has remained an open question. Using computer-aided algorithm analysis, we show that the gradient-evaluation sequence also achieves the optimal order of iteration complexity in this constrained setting, thereby extending the theoretical understanding of accelerated methods for constrained optimization.

26. Association of Sociodemographic Factors with Referral to Specialized Eating Disorder Care Before and After the 2023 Guideline Update

Salma Akter, Dr. Bing Zhang Department of Statistics, University of Kentucky

Abstract: Feeding and eating disorders (FEDs), including anorexia nervosa, bulimia nervosa, and binge eating disorder, are serious conditions affecting adolescents and young adults and remain underdiagnosed despite increasing prevalence. In February 2023, the American Psychiatric Association (APA) updated clinical guidelines to emphasize earlier recognition and multidisciplinary referral; however, the impact of these changes on referral patterns is not well characterized. This study evaluates changes in referral to specialized eating disorder care before and after the guideline update and examines associations with sociodemographic factors. We conducted a retrospective cohort study using de-identified electronic health record data from patients aged 13–25 diagnosed between February 2021 and February 2025. Patients were classified into pre- and post-February 2023 cohorts. The primary outcome was placement of at least one specialty referral. Logistic regression models were used to estimate odds ratios (ORs) and 95% confidence intervals (CIs). Referral rates increased from 60.4% in the pre-guideline cohort to 77.1% in the post-guideline cohort, with higher odds of referral following the update (OR = 2.20, 95% CI: 1.48–3.31, p ¡ 0.0001). Geographic location was not associated with referral likelihood (OR = 0.91, 95% CI: 0.61–1.37). Increasing age was associated with lower referral probability, and referral likelihood differed by insurance type. These findings suggest that guideline implementation improved referral rates, though disparities persist across sociodemographic groups.

27. Empirical likelihood inference for the Location-Adaptive Change-Point Testing for Time Series

Ali Jinnah, Mathematics and Statistics, Georgia State University

Abstract: We propose a nonparametric change-point detection procedure for time series by incorporating empirical likelihood (EL) into the location-adaptive self-normalization (SN) framework of Dai and She (2021). Classical SN-based CUSUM tests are known to suffer power loss when change points occur near the boundaries of the observation window. While Dai and She (2021) address this through an optimal location-adaptive subsampling scheme, the empirical likelihood methodology of Wang and Ning (2025) offers a nonparametric estimation approach with favorable properties under small training samples and unknown distributional settings. We extend the location-adaptive SN framework by incorporating the EL-based method of Wang and Ning (2025), with the goal of improving detection performance of Dai and She's procedure. We conduct simulation studies to evaluate the performance of the proposed approach and assess its improvement upon the location-adaptive SN method of Dai and She (2021). A well-log data analysis further illustrates the practical utility of the method.

28. **Causal inference on multiple exposures in the presence of invalid instruments**
Jiwon Kang, Statistics, Florida State University

Abstract: Instrumental variable methods are fundamental for causal inference with endogenous exposures, but their validity depends on assumptions that are often difficult to verify in practice. This challenge is especially acute in multivariable Mendelian randomization, where correlated exposures are analyzed jointly and invalid instruments can induce structured bias across effect estimates. We study identification and estimation of multivariable causal effects in linear models with individual-level data and potentially invalid instruments. We establish identifiability under a multivariable extension of plurality-based identification and propose a two-stage hard-thresholding procedure that separates joint relevance screening from validity assessment via pairwise compatibility and voting. We prove selection consistency for the estimated valid instrument set and asymptotic normality for the resulting two-stage least squares estimator. Simulations show near-oracle performance across diverse regimes, including settings in which most relevant instruments are invalid. In an analysis of Alzheimer's Disease Neuroimaging Initiative (ADNI) data, the method yields more stable and interpretable conditional region-specific effects by attenuating signals driven by correlated structures in single-exposure analyses.

29. **Statistical Methods for Identifying Predictive Biomarkers Using Time-To-Event Ratios**
Zhimi Kuang, Statistics, University of Kentucky

Abstract: Despite the development of novel cancer therapies, treatment effects remain heterogeneous among patients. Therefore, identifying predictive biomarkers capable of stratifying patients into subgroups with differential treatment benefits is essential. This is particularly important in early-phase, non-randomized phase II trials, which can inform the design of subsequent randomized phase II or biomarker-enriched phase III trials. However, the absence of a control arm in non-randomized trials poses a substantial challenge in distinguishing predictive biomarkers from prognostic biomarkers. The Progression-Free Survival (PFS) ratio design - or more generally, the Time-to-Event (TTE) ratio design - offers a promising framework for addressing this issue. We propose a semiparametric accelerated failure time (AFT) model based on TTE ratios to identify predictive biomarkers. Our model is flexible and yields parameters with direct clinical interpretation. We further develop a weighted kernel smoothed log-likelihood estimation method for the model, accounting for the dependent censoring of TTE ratios. In contrast, existing parametric or nonparametric AFT estimation methods require independent censoring assumption and may suffer from substantial estimation bias for the TTE ratio data. Extensive simulation studies demonstrate that our method performs well in practical settings and outperforms existing methods. We illustrate the utility of our method using data from the WINTHER trial, which employs the PFS ratio as its primary endpoint.

30. **Random graph asymptotics for treatment effect estimation under infinite rank graphon model**
Nirmalya Mandal, Department of Statistics, North Carolina State University

Abstract: Network data arises naturally in fields such as sociology, marketing, and medicine. Quantifying causal effects of treatment assignments is increasingly important when experimental units interact. Causal inference under network interference is challenging due to violations of the Stable Unit Treatment Value Assumption (SUTVA) underlying the Neyman–Rubin potential outcomes framework. Most existing work studies this problem under a fixed exposure graph. Recently, Li and Wager (2023) considered settings where the exposure graph is generated from a graphon model, and studied estimation of direct effects (effects of a unit's own treatment) and indirect effects (effects of neighbors' treatments). They proposed a PC-balancing estimator for the indirect effect and established large-sample asymptotic results under a fixed-rank graphon assumption. In this work, we relax this assumption by allowing the rank of the graphon to grow with the sample size. We establish analogous large-sample guarantees and characterize the trade-off between graphon approximation error and the rate of asymptotic normality.

31. **Multinomial Logistic Modeling of Mental Health Outcomes: Impact of Dysmetabolic Syndrome Burden in Head and Neck Cancer Survivors**
Shrishti Sharma, Mathematics and Statistics, University of West Florida

Abstract: Approximately 70% of cancer survivors have at least one chronic comorbidity, increasing the risk for mental health disorders (MHDs). However, the joint relationship between dysmetabolic syndrome burden (MetS) and distinct mental health outcomes remains unclear. Modeling co-occurring outcomes such as depression and anxiety therefore requires approaches that extend beyond binary classification.

We analyzed electronic health record data from the NIH All of Us Research Program, identifying 1,653 head and neck cancer (HNC) survivors using ICD-9/10 codes. MHDs were defined as a four-category outcome: none, depression only, anxiety only, or comorbid depression and anxiety. MetS burden was quantified as the count of five components (obesity, diabetes, hypertension, low HDL, and high triglycerides) and categorized as 0, 1, 2, 3, or ≥ 4. A multinomial logistic regression model was used to estimate relative risk ratios (RRRs) for each outcome category relative to the reference group (no MHD), adjusting for demographic, behavioral, and treatment-related covariates.

The study population was predominantly male (62.01%) and White (69.93%), with a mean age of 69.64 years (SD = 11.77 years). Increasing MetS burden was associated with higher RRRs across all MHD categories, demonstrating a clear dose–response relationship. Compared to individuals with no MetS components, higher MetS counts were associated with increased RRRs for depression only (ranging from 1.91 to 4.54), anxiety only (ranging from 2.15 to 3.24), and comorbid depression and anxiety (ranging from 2.20 to 13.58). The strongest associations were observed for the comorbid outcome; individuals with ≥ 4 MetS components had a significantly elevated RRR (13.58; 95% CI: 8.09–22.8). Older age and male sex were associated with lower RRRs, whereas unmarried status and smoking history were associated

with higher RRRs. Cancer treatment burden showed modest associations, with the receipt of chemotherapy, radiation, and surgery linked to an increased RRR for comorbidity.

Among HNC survivors, increasing MetS burden is strongly associated with adverse mental health outcomes, particularly comorbid depression and anxiety. These findings highlight the importance of modeling multi-category outcomes and support integrated approaches to survivorship care.

32. Two-Stage Gene–Environment Interaction Analysis via Forest-Based Screening and Sparse Interpretable Modeling

Kakon Datta, Dr. Bing Zhang Department of Statistics, University of Kentucky

Abstract: Detecting gene–environment (G×E) interactions is challenging because interaction effects are typically weak, genetic data are high-dimensional, and environmental effects may be complex and heterogeneous. We propose a two-stage framework that combines nonparametric screening with sparse, interpretable modeling for high-dimensional G×E analysis. First, SNPs are aggregated within gene-level sets, and a random forest is used to screen and rank candidate gene–environment interactions. Second, the retained candidates are analyzed with a sparse regression model including both main effects and interaction terms, producing effect estimates and an interpretable interaction structure. To improve stability, we incorporate uncertainty-aware interaction importance based on model-averaged inclusion across candidate sparse models. The method is motivated by the sparsity of true interaction signals and the need to reduce the large search space before formal inference. Preliminary investigations indicate stable performance under diverse G×E settings, and further results from simulation studies and a depression-related application will be presented.

33. Quantifying Predictive Uncertainty in Student Academic Outcomes: A Conformal Prediction Framework

Kazi Sabbir Ahmad Nahin, Dr. Bing Zhang Department of Statistics, University of Kentucky

Abstract: Universities can provide more timely and effective support when students at risk of dropout are identified early. However, most machine learning (ML) models used for this purpose produce only a single predicted label (e.g., “graduate” or “dropout”) without indicating the confidence associated with that prediction, limiting their usefulness in high-stakes intervention settings. In this study, we employ Inductive Conformal Prediction (ICP) to convert point predictions into prediction sets with user-defined coverage guarantees. We also examine an ensemble extension that combines conformal p-values using Fisher’s method, allowing us to evaluate how ensemble composition affects predictive validity and efficiency. The framework is assessed using both a multi-scenario simulation study and a real-world higher education dataset, considering binary (Graduate vs. Dropout) and three-class (Graduate, Enrolled, Dropout) formulations. In simulation, individual ICP models remain close to nominal targets across varying class imbalances and task complexities. At 95% confidence, empirical coverage ranges from 0.92–0.97 (binary) and 0.94–0.98 (multiclass), with efficient prediction set sizes of 1.07–1.85 and 1.73–2.59, respectively. These trends are consistent with real-data results.

While baseline ML accuracy is higher for the binary task (˜0.89–0.91) than the multiclass task (˜0.63–0.78), ICP provides calibrated uncertainty, producing compact sets in the binary case and larger, more cautious sets in the multiclass setting. Fisher-based ensemble conformal prediction is found to be sensitive to dependence among base learners. Ensembles built from highly similar models show under coverage (0.86–0.89 at 95% confidence) and occasional empty prediction sets, while more diverse ensembles improve stability. Overall, conformal prediction offers a practical framework for uncertainty-aware student outcome forecasting, with model diversity playing a key role in ensemble reliability.

34. Reliable score-based confidence intervals for covariance parameters in linear mixed models

Matias Shedden, Statistics, University of Florida

Abstract: Uniform inference refers to the process of constructing confidence intervals or hypothesis tests which are uniformly asymptotically valid on compact subsets of the parameter space. Uniform results generally follow from distributional results established under a sequence of parameter values θ_n. We establish such distributional results for the profile score statistic, evaluated at an unrestricted maximum of the nuisance parameters. We then consider confidence intervals constructed by inverting the profile score statistic, and show that these confidence intervals can be shown to be uniformly asymptotically valid under certain conditions, which we discuss in the case of Linear Mixed Effects models. In particular, our proposed confidence intervals are valid even when the true value of the parameters, including the nuisance parameters, may be at or near the boundary of the parameter set. This provides a reliable method for constructing confidence intervals for a single parameter value, and is particularly valuable when estimates for the parameter values are highly correlated, and we discuss such applications.

35. Denoising Coupled Maximum Entropy Estimation for Relaxation-Diffusion Distribution of dMRI

Xunan Yang, Department of Statistics, University of South Carolina

Abstract: We consider the inverse problem of estimating a collection of spatially indexed density functions from relaxation-diffusion MRI data. The measured signal is related to the unknown distributions through a multidimensional integral transform, yielding a high-dimensional and ill-posed inverse problem with both functional and spatial structure. To address the scalability limitations of classical discretization approaches such as basis expansion or dictionaries, we develop a maximum-entropy framework. The proposed method represents the target density as a base distribution modulated by an exponential tilt, where the base distribution may be either uninformative or informative, for example learned using deep generative models. Regularization on signal space is incorporated via denoising algorithms like total variation. A highly scalable alternating direction method of multipliers algorithm is derived for point estimation. Numerical experiments demonstrate that the proposed method yields improved density estimates compared to standard alternatives.

36. Survey-Weighted Logistic Regression of Metabolic Syndrome Components and Depression in Reproductive-Aged Women: NHANES 2011-2023

Shelby Blair, Public Health, University of West Florida

Abstract: Introduction: Depression and metabolic syndrome are interconnected public health concerns, particularly among women of reproductive age. Prior literature has shown that metabolic syndrome has a bidirectional relationship with depression. This study examined depression and metabolic syndrome using nationally representative data from a complex survey model. Methods: This study used National Health and Nutrition Examination Survey (NHANES) data from 2011-2012 through 2021-2023. The analytic sample included women aged 15-49 with complete data on depression and metabolic syndrome components. Survey-weighted descriptive and bivariate analyses were conducted. Associations were estimated using survey-weighted logistic regression models accounting for stratification, clustering, and sampling weights. Models were adjusted for age, race/ethnicity, education, and marital status. Results: The weighted sample represented 44.81 million U.S. women with a mean age of 34.09 years (SE=0.22). Logistic regression modeling demonstrated that increasing metabolic syndrome components were significantly associated with higher odds of depression, with the highest odds observed among individuals with two components (OR = 2.80, 95% CI: 1.52–5.17), followed by those with three or more components (OR = 2.61, 95% CI: 1.20–5.68). Education was associated with depression, including those with a high school education or less (OR = 1.81, 95% CI: 1.18–2.78, p = 0.01) and those with some college education (OR = 1.79, 95% CI: 1.10–2.92, p = 0.02). Conclusion: These findings highlight depression and metabolic health using survey-weighted modeling approaches that account for complex sampling design. This highlights the importance of using survey-weighted approaches that account for complex sampling structures that incorporate sampling weights, clustering, and stratification to produce population-representative estimates.

37. Procrustes Alignment in Latent Distance Models for Bipartite Data

Brian Harrold, Department of Statistics, University of South Carolina

Abstract: For bipartite data, consider two distinct types of observations whose positions in a shared latent space are being estimated. The model is defined through distances calculated between pairs with one member from each group. When these locations are estimated using Markov chain Monte Carlo (MCMC) methods, the resulting configurations are non-identifiable, as distances are invariant to rotation, reflection, and translation. Consequently, latent positions may vary substantially across posterior draws, even when the underlying distance structure has converged. Procrustes alignment is commonly used to address this issue by transforming each draw to a common reference configuration. However, limited practical guidance exists for implementing Procrustes alignment in bipartite settings. In particular, questions arise regarding whether alignment should be performed jointly or separately across partitions. Alternative Procrustes alignment strategies are evaluated for latent distance models applied to bipartite data. Simulation studies assess how these approaches influence the stability of latent configurations, recovery of distance structure, and interpretability of the resulting embeddings. The results provide practical recommendations for alignment and highlight the role of these choices in shaping conclusions.

38. Quantifying the Association Between Prior-Season Performance and Salary in the NBA

Mohammad Kamrul Alam, Dr. Bing Zhang Department of Statistics, University of kentucky

Abstract: Player salary structures in the National Basketball Association reflect both individual contributions and team-level decision making processes. This study investigates how player performance in the 2022–23 season is associated with salary outcomes in 2024. Player salary data were linked with prior-season performance statistics using unique identifiers, and a log transformation was applied to salary to address strong right-skewness.We fitted a multiple linear regression model including points per game (PTS), assists (AST), rebounds (TRB), minutes played (MP), three-point percentage (3P%), and age as explanatory variables. The model accounted for a substantial proportion of variation in salary (adjusted R2=0.613, p¡0.001). Scoring output, playing time, and age emerged as the most influential predictors, with estimated increases in salary of approximately 5.5%, 4.5%, and 4.4% per unit increase, respectively. In contrast, assists, rebounds, and three-point efficiency did not show statistically significant associations after controlling for other variables.These results suggest that salary allocation in the NBA is primarily driven by measures of player usage and role, rather than all aspects of on-court performance. The findings provide a quantitative perspective on how performance metrics translate into compensation and highlight the relative importance of key indicators in shaping salary outcomes.

39. Psychometric Validation of the Short-Form Medical Outcomes Social Support Scale Using Exploratory Factor Analysis and Reliability Assessment

Erin Wade, Public Health, University of West Florida

Abstract: Background: The Medical Outcomes Social Support Scale (MOS-SSS) is a 19-item tool scored on a 5-point Likert Scale developed in late 1980 to screen for social support in the general population. No study has validated its use among college going youth. The aim of this study was to evaluate the short-form version using statistical methods such as measures of reliability and construct validity. Methods: We modified the 19-item MOS-SSS scale to 7-items based expert reviewers. This 7-item MOS-SSS short form was applied to 119 participants at a public university. Analysis was conducted via content validity using 2 expert raters, reliability analysis using Cronbach's alpha, concurrent validity against Patient Health Questionnaire-9 (PHQ-9) and Generalized Anxiety Disorder-7 (GAD-7) and construct validity through exploratory factor analysis (EFA) with principal component extraction and Oblimin rotation with Kaiser normalization. Results: The sample was predominately white (74.0%), female (79.9%), unmarried (73.9%), pursuing a master's degree (47.6%) with a mean age of 28.8years (SD = 10.5years). Content validity was strong and reliability was confirmed using Cronbach' s α (0.817) after item refinement. Concurrent validity was strong (r= -0.84, p¡.001) EFA explained 70.23% variance and identified a 2-factor solution for 6 items: tangible support and positive social and emotional interaction. Conclusion: The results provide satisfactory evidence of reliability and construct validity with the identified factor solution supporting the scale's latent structure. These findings suggest that the short-form is statistically adequate in identifying social support among college students, however further validation in larger samples is warranted.

40. Adaptive Task Grouping for Multi-Task Learning in High-Dimensional Time Series

Lu Zhou, Department of Statistics, University of South Carolina

Abstract: Multi-task learning (MTL) can improve statistical efficiency in high-dimensional, irregularly sampled time series by borrowing strength across related tasks, but its effectiveness depends critically on task relatedness and the structure of shared signal. This project develops a unified framework for adaptive MTL, motivated by electronic health record (EHR) data, that combines gradient-based task grouping with flexible deep sequence architectures. We propose a two-stage approach in which gradient-based similarity measures are first used to identify groups of related tasks, after which tasks within each group are modeled jointly using GRU-decay recurrent networks with transformer-based cross-task attention and dynamic task weighting to address missingness, imbalance, and heterogeneous signal strength. Controlled simulations systematically vary shared signal strength, nonlinearity, missingness, and sample size to isolate their effects on task grouping and predictive performance. The results provide methodological insight and practical guidance on when adaptive task grouping yields positive transfer in high-dimensional time series settings.

41. Classification of Separable Nonlinear Signals: Performance Limits

Pedro Izquierdo Lehmann, Applied Mathematics and Statistics, Johns Hopkins University

Abstract: This paper considers the statistical problem of classifying a signal corrupted by noise. We focus on separable nonlinear signals: linear combinations of functions that depend nonlinearly on the problem's parameters. We assess how the Bayes error of the classification task worsens due to noise, considering the regularity of the data's statistical distribution and the structure of the separable nonlinear signal. Typically, the noise level and the signal's structure depend on the measuring instrument used. Therefore, our analysis provides guidance on prioritizing error mitigation when developing new measurement devices. We demonstrate our findings by analyzing two cases: signals as sums of complex exponentials, relevant to nuclear magnetic resonance data, and signals as sums of Gaussians, relevant to microscopy.

42. Doubly Robust Angle based Direct Learning for Optimal Individualized Treatment Rules using AIPW Pseudo Outcomes with Super Learner Estimation

Apsara Pitigalaarachchi, Bioinformatics & Biostatistics, University of Louisville

Abstract: Since every patient reacts differently, treatment effect heterogeneity (TEH), which describes variations in treatment responses between individuals, are becoming increasingly crucial in clinical research, because it sheds light on developing precision medicine. Conditional Average Treatment Effects (CATE), the difference in the conditional mean outcome between treatments given covariates, is crucial to study TEH and identify individualized treatment rules (ITR) to maximize patient outcomes. In practice, estimating an optimal ITR that maximizes a patient's anticipated clinical outcome based on the patient's characteristics is a significant challenge. The existing methods are often computationally expensive and/or not very accurate. In this work, we propose a technique combining an angle-based strategy for

multiple treatments with doubly robust pseudo-outcomes to accurately predict the optimal ITR. We employ the SuperLearner to implement an ensemble of models to prevent model misspecification and achieve accuracy, while simultaneously estimating all possible pairwise CATE to improve efficiency. We conduct numerical studies to validate the merits of the proposed method.

Keywords: Treatment Effect Heterogeneity, Conditional Average Treatment Effects, Optimal Individualized Treatment Rule, SuperLearner

43. Registration-Free Shape Analysis of Decedent Organs via Topological Data Analysis

Yanyan Zhan, Department of Statistics, University of South Carolina

Abstract: Quantifying postmortem decomposition from organ shape is challenging because standard statistical shape analysis relies on well-defined pointwise correspondence, an assumption that becomes ill-posed for volumetric data with complex and variable internal structure. This problem is especially pronounced in forensic CT data, where organs may exhibit fragmentation, internal voids, and strong heterogeneity across samples. In this work, we study brains and kidneys segmented from forensic CT scans and represent each organ as a 3D volumetric point cloud to capture both external morphology and decomposition-induced interior structure. Registration-dependent methods, including principal component analysis and principal nested cones, show limited ability to characterize variation, especially for highly degraded or irregular shapes. To overcome this limitation, we develop a registration-free framework based on topological data analysis. Persistent homology, computed via alpha-complex filtration of the full point clouds for each organ, is summarized using scalar, functional, and distance-based representations that capture complementary aspects of decay, including cavity formation, fragmentation, and multiscale structural irregularity. These topological features are integrated to cluster organs into decomposition stage groups and to construct decomposition atlases ordered along interpretable decay trajectories. The results reveal distinct, organ-specific patterns of degradation and show that brain and kidney decomposition can progress asynchronously within the same individual. More broadly, the proposed framework provides an interpretable and robust approach for analyzing volumetric shapes with complex internal structure, where correspondence is ambiguous or undefined.

44. Opportunistic asthma self-management education in routine care improves lung function in children

Brooke Nash, Public Health, University of West Flordia

Abstract: Background: Asthma affects 6.5% of American children and contributes substantially to emergency visits and unplanned hospitalizations. While asthma self-management education (AS-ME) improves outcomes, most evidence comes from structured programs. Less is known about the effectiveness of flexible, clinic-integrated education delivered opportunistically by community health workers (CHWs). This study evaluated whether this model is associated with improved lung function in pediatric patients with asthma.

Methods: A retrospective chart review was conducted for pediatric patients aged 2–17 years with asthma between June 1, 2021, and June 30, 2024. Patients with a single visit or ¡6 months in the practice were excluded. A CHW delivered brief, patient-tailored education during routine visits, including adherence support and reinforcement of provider care plans. 101 children with 313 spirometry-recorded visits were analyzed. The primary outcome was FEV$_1$%. Linear mixed-effects models with random intercepts assessed longitudinal associations, adjusting for demographic and clinical covariates.

Results: The cohort was predominantly male (74%) and non-Hispanic Black (33%). Baseline lung function was similar between children with and without prior AS-ME. In linear mixed-effects models with random intercepts identified strong within-child dependence and substantial baseline effects. Baseline FEV$_1$% predicted was a major determinant of follow-up values (β = 0.47, 95% CI: 0.30–0.64; p ¡ .001), while oral corticosteroid use was associated with significantly lower lung function (β = −11.34, 95% CI: −21.47 to −1.20; p = .029). The cumulative ASME × visit interaction was statistically significant (F(3,164.28) = 4.54; p = .004), indicating differential longitudinal slopes. Children with two prior ASME sessions showed the steepest improvement (β = 7.99 per visit, 95% CI: 3.71–12.27; p ¡ .001), with smaller positive slopes for one session (β = 4.15; p = .042) and $\geq$3 sessions (β = 4.42; p = .055). Sensitivity analyses excluding baseline FEV$_1$ adjustment produced consistent interaction effects.

Conclusion: Opportunistic asthma education delivered by an embedded CHW as part of routine pediatric care was associated with improved lung function trajectories. This model may offer a scalable approach to improving asthma outcomes.

45. Sparse Longitudinal Functional Principal Components Analysis

Nidhi Pai, Division of Biostatistics and Health Data Science, University of Minnesota

Abstract: We propose sparse longitudinal functional principal components analysis (sparse LFPCA) for analyzing sparsely or irregularly sampled functional data collected repeatedly for each subject. While existing methods for LFPCA are limited to densely observed curves, we accommodate sparse data by casting covariance estimation as a structured penalized spline regression problem. This framework enables simultaneous estimation and smoothing of multiple covariance components while borrowing information across locations in the functional domain. Extensive simulations show that sparse LFPCA (1) provides accurate estimates of eigenfunctions and generates reasonable predictions for underlying curves from sparse longitudinal functional data, and (2) achieves similar or superior performance compared to existing methods for dense longitudinal and sparse multilevel functional data. Methods are motivated by and applied to the Intern Health Study to investigate smartphone typing speed as a potential scalable ambulatory assessment of mental fatigue among first-year training physicians in the US. Sparse LPFCA reveals interpretable participant- and day-level patterns not captured by a previous analysis.

46. Concentration of Sparse Random Graphs

William Beatty, Applied Mathematics and Statistics, Johns Hopkins University

Abstract: Sparse random graphs whose expected degrees are $o(\log n)$ do not concentrate close to their expectations because of a small number of vertices with abnormally high or low degree. Existing work shows that reweighting edges to compress the degree distribution results in concentration, but these results exclude some models of random graph with low expected degree. We present ongoing work to generalize these results to a wider class of sparse random graphs.

47. Query Complexity of Black-Box Model Classification

Merrick Ohata, Applied Mathematics and Statistics, Johns Hopkins University

Abstract: We consider the problem of classifying a black-box generative model based on its responses to a collection of queries. Not all query sets are equally discriminative; certain queries produce strong class separation while others do not. We formalize this distinction through the discriminative factorization, a decomposition of query-model interaction into independent statistical ”directions” to separate informative and uninformative queries. Under this framework, the probability of chance-or-worse classification decays exponentially in the query budget at a rate governed by parameters that can be estimated from the spectral structure of a query-model matrix. We demonstrate that the estimated parameters predict the empirical decay rate, and query sets selected using the estimated discriminative factorization reproduce oracle query selection, thereby improving classification efficiency without requiring task-specific knowledge.

48. Trial-Level Exploratory Analysis of Clinical Trial Non-Completion in the United States: A 10-Year Retrospective Analysis of AACT Registry Data

Emmanuel Paalam, Mathematics and Statistics, University of West Florida

Abstract: Background: Clinical trial non-completion remains a major barrier to efficient evidence generation, with reported rates of 20–30%. However, recent large-scale analyses incorporating rigorous data preprocessing are limited. Objective: To quantify trial non-completion and identify trial-level characteristics associated with completion outcomes. Methods: We conducted a retrospective analysis of 85,414 clinical trials from ClinicalTrials.gov (AACT), including U.S.-based interventional and observational studies initiated between 2015 and 2024. Extensive data cleaning and wrangling were performed, including harmonization of trial status definitions, standardization of sponsor and phase variables, removal of duplicate and incomplete records, and derivation of study duration. Trials were classified as completed, not completed (terminated, withdrawn, suspended, or unknown), or in progress. Predictors included U.S. facility presence, sponsor type, study type, trial phase, and study duration. Analyses were performed in R (v4.3.2) using nonparametric methods (χ^2 and Kruskal–Wallis tests with Bonferroni-adjusted Wilcoxon comparisons). Results: A total of 16,159 trials (18.9%) were not completed. All predictors were significantly associated with trial status (p ¡ 0.05). Non-completed trials were more likely to lack domestic sponsorship (5% vs 1%), be interventional (87% vs 84%), and be phase II studies (36% vs 30%) compared with completed trials. Conclusions: Trial non-completion remains substantial and is associated with identifiable structural characteristics. Rigorous data preprocessing is critical for valid inference in registry-based research. Targeted interventions focusing on sponsorship, trial design, and mid-phase studies may improve completion rates.

49. The price of service: A comparative study of money and military retention using time series and predictive analytics

Yvena Aristilde, Mathematics, University of West Florida

Abstract: The ability of the United States military to maintain a strong and ready force depends largely on retaining its most experienced personnel. However, predicting why service members choose to stay or leave remains a significant challenge for policymakers. This thesis examines effects of defense spending and regular military compensation influence on retention, with a specific focus on the differential responses of officers and enlisted personnel respond to financial incentives. The research employs two complementary approaches to investigate potential solutions. First, vector autoregression (VAR) models are used to capture dynamic relationships in historical data and evaluate how changes in military pay influence force size over time. Second, long short term memory (LSTM) network, a type of recurrent neural network (RNN), is used to model sequential patterns and forecast differential responses of officers and enlisted personnel response to financial incentives. Both models suggested that the monetary incentives differ among officers and enlisted personnel. Officer personnel respond to greater sensitivity to immediate pay raises and bonuses, whereas enlisted'response to a broader set of long-term career incentives and overall defense spending. By understanding these differences, the military can move away from standardized policies and develop rank-based strategies that more effectively retain highly skilled personnel.

50. From Association to Causal Inference: The Impact of Neighborhood Environments on Depressive Symptoms among Older Adults in U.S.

Abida Sultana Asha, Statistics, University of Kentucky

Abstract: Background: As the U.S. population ages, preserving mental health in later life has become a critical public health priority. Depressive symptoms among older adults act as catalysts for physical and cognitive decline. While previous research links neighborhood environments to psychological well-being, majority of this literature is strictly associational, making it difficult to unravel true environmental effects. For this reason, understanding the causal effect of social and environmental conditions on depressive symptoms among older adults is an important area of inquiry. Objective: This study investigates both the associational and causal effects of perceived neighborhood social cohesion and physical disorder on depressive symptom burden among adults aged 50 and older. Methods: Data were drawn from the 2020 Health and Retirement Study (HRS) psychosocial subsample (N=4,328). The outcome was the 8-item CES-D scale. The study proceeded in two phases. Phase I utilized survey-weighted negative binomial regression to establish baseline associations, adjusting for age, gender, education, transformed household wealth, and self-rated health. Under the causal inference framework in Phase II, where we implemented a Doubly Robust (DR) Augmented Inverse

Probability Weighting (AIPW) estimator for continuous exposures. Generalized Propensity Scores were calculated using Covariate Balancing Propensity Score (CBPS) optimization, combined with the Parametric G-Formula. Non-parametric bootstrapping generated Bias-Corrected and Accelerated (BCa) 95% confidence intervals. Results: Phase I revealed significant associational incidence rate ratios for both social cohesion (IRR=0.82) and physical disorder (IRR=1.15). Phase II successfully balanced all confounders and isolated the causal Average Dose-Response Functions. The DR estimator yielded an Average Marginal Effect (AME) of -0.181 (95% BCa CI: -0.238, -0.126) for social cohesion, and an AME of 0.144 (95% BCa CI: 0.089, 0.195) for physical disorder. Conclusion: Physical disorder actively causes psychological distress in older adults, while social cohesion causally mitigates it. These results underscore the necessity of integrating neighborhood-level maintenance and social infrastructure into broader public health strategies to effectively protect the mental well-being of the aging population.

51. Bayesian Spatial Additive Regression Using Lancaster–Salkauskas Natural Cubic Spline Basis Functions and Matérn Gaussian Processes

Mengyan Jing, Department of Statistics and Data Science, University of Missouri

Abstract: Spatially referenced data with nonlinear covariate effects arise in many applied settings, including environmental science, economics, and public health. Existing approaches such as Bayesian P-splines with Markov random field spatial priors (Lang and Brezger, 2004) rely on B-spline bases with different penalties and discrete spatial structures that may not reflect the continuous nature of the underlying spatial process. We propose a Bayesian spatial additive model that pairs Lancaster–Salkauskas (LS) natural cubic spline basis functions with a continuous Matérn Gaussian process spatial random effect. The LS basis parameterizes spline coefficients directly as function values at the knots, yielding interpretable parameters and enabling principled prior specification—advantages not shared by B-spline coefficient representations. Inference proceeds via a Markov chain Monte Carlo sampler that analytically marginalizes the spatial random effects, avoiding a variance–spatial coupling instability that arises when the spatial effects are sampled explicitly. Smoothing variances and spatial dependence parameters are updated through Metropolis–Hastings steps within the Gibbs sweep, with REML estimates providing starting values for the chain.

We evaluate the method through simulation studies examining scenarios with and without spatial dependence, varying nonlinear signal complexity, and different sample sizes. Results demonstrate accurate nonlinear function recovery, well-calibrated 95% credible intervals, and robustness across a range of prior specifications. We apply the model to Munich rent data, recovering nonlinear effects of floor space, year of construction, and other housing characteristics while accounting for spatial dependence across city districts, and compare our findings to the benchmark analysis of Lang and Brezger (2004). Extensions to count outcomes via a Poisson log-link formulation are currently under development.

Boyd Harshbarger, PhD (Feb. 15, 1906 – May 20, 1998)

Boyd Harshbarger was one of the early pioneers of statistics in the United States. He received his bachelor's degree from Bridgewater College, and his M.S. in Mathematics from the University of Illinois, and an M.S. from Virginia Tech. His career at Virginia Polytechnic Institute began in 1931, where he founded one of the earliest (third oldest) Departments of Statistics in the country. In 1935, he organized and taught the first courses in statistics in the Department of Mathematics. He received a Rockefeller Fellowship and left to pursue his Ph.D. in 1940. He wrote his doctoral dissertation under the direction of distinguished Professor W.G. Cochran at Iowa State College and George Washington University, and returned to VPI in 1942.

The success of the VPI Department of Statistics was due to the foundation laid by Boyd Harshbarger. Boyd Harshbarger was the founder of the Statistical Laboratory at VPI in 1947 and the Department of Statistics (1949). He was the head of the Department of Statistics from 1949 to 1972, and was named professor emeritus in 1976.

He received many honors in recognition of his services to the statistics profession and to the sciences in general. In addition to receiving an honorary doctorate from his alma mater, he was named a fellow of the American Association for the Advancement of Science (AAAS), the American Statistical Association (ASA), the Institute of Mathematical Statistics (IMS), and the Virginia Academy of Science. He was a charter member of the Biometric Society and president of the Eastern North American Region (ENAR) from 1956–1958. The VAS honored his work with the J. Shelton Horsley Research Award (1946), with the Ivey F. Lewis Distinguished Service Award (1966), and with honorary life membership. Harshbarger organized the Statistics Section of the VAS, and served as its secretary (1945–1946), vice-chair (1947) and chair (1948).

With respect to today's Southern Regional Council on Statistics, a Committee on Statistics was coordinated through the Southern Regional Education Board, and consisted of representatives from 14 southern states, from Maryland to Texas. (This was the precursor to today's SRCOS.) Their charge was to promote statistics graduate programs in the southern colleges and universities. Boyd Harshbarger was the first chairman of this committee (1961–1962). In 1954, Harshbarger organized and supervised the first Regional Statistical Summer Season (the first "Summer Research Conference") at VPI, bringing together scientists from five continents and students from 35 states of the US. Harshbarger influenced many students and colleagues with his enthusiasm and vision of statistics. The Boyd Harshbarger travel awards were giving in 1994 at the SRC in Williamsburg, Virginia, supported entirely by the SRCOS. Beginning with the 2000 SRC, also in Williamsburg, support in the form of grants from federal agencies was obtained for this student award, along with the R. L. Anderson Award.

Richard L. Anderson, PhD (April 20, 1915 – Jan. 19, 2003)

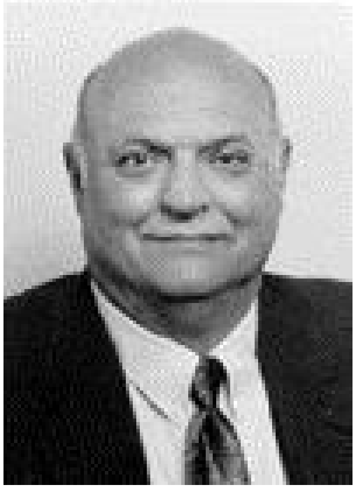
Richard "Dick" L. Anderson lived through the farm depression of the 1920s and the general depression of the 1930s. He missed two elementary grades, so that he was two years younger than his high school class mates. Income from clover seed bags was used to pay for his room and board at Depauw University, where he had a tuition scholarship, and he eventually went on to earn the AB degree in 1936, and his Ph.D. in Mathematics, Statistics, and Economics at Iowa State College in 1941. He began his academic career at North Carolina State College and rose from Instructor to full professor in his 25 years of service there to 1966. He was away from Raleigh in 1944–45 when he consulted with the Army and Navy at Princeton, in 1950–51 when he visited Purdue, and during 1958 when he spent time at the London School of Economics. In 1967, after a year as a visiting research professor at the University of Georgia, he took up the chairmanship of the newly established Department of Statistics at the University of Kentucky.

Professor Anderson had long-standing interests in experimental design, regression methods, variance components and time series analysis and their application to agricultural, industrial and operational problems. He took an active role in statistics in the US and internationally. He was a fellow of the American Statistical Association (ASA), the Institute of Mathematical Statistics, and of the American Association for the Advancement of Science. He was President of the Eastern North American Region of the Biometric Society, was a member of the International Statistical Institute, and was President of the ASA. He was a joint author of the well-known book *Statistical Theory in Research*.

Dick had a long association with the American Statistical Association, and was elected as a fellow in 1951, president in 1982, and received a Founder's Award in 1992. He was also a fellow of the American Association for the Advancement of Science and the Institute for Mathematical Statistics. He was a member of the International Statistical Institute, and the International Biometric Society.

Dick Anderson was one of the panel of original organizers of the Regional Statistical Summer Season (the precursor to the Summer Research Conference) and was a long-time member of SRCOS, and served as president of SREB-COS from 1975–1977. In recognition of his many contributions to the advancement of statistics in the South and nationally, graduate student travel awards were made annually in honor of Richard L. Anderson. The first awards were giving in 1994 at the SRC in Williamsburg, Virginia, supported entirely by the SRCOS. Beginning with the 2000 SRC, support in the form of grants from federal agencies was obtained for this student award, along with the Boyd Harshbarger Travel Awards.

M. Clinton Miller, PhD (Aug. 28, 1932 – Feb. 10, 2004)

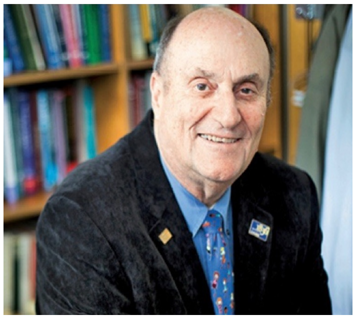


Dr. Miller received his high school diploma at Oklahoma Military Academy, where he also began his college work. He graduated from the University of Oklahoma with a degree in chemistry, and completed the first year of medical school at the Oklahoma University School of Medicine. His medical career goal was temporarily interrupted by a diving accident that left him a quadriplegic at the age of 21. Combating difficult odds, he soon returned to the University of Oklahoma, where he earned a master's degree in mathematics and to the Oklahoma Medical School, where he received a doctorate in biostatistics. He then did a number of post-doctoral fellowships and seminars.

Dr. Miller's long and distinguished career included teaching and research at the Medical Schools of Oklahoma, Tulane University in New Orleans, and the Medical University of South Carolina, where he established the Department of Biometry. During his 26 year tenure, he authored or co-authored 19 books and wrote or contributed to more than 81 scientific articles and publications. He worked on committees and projects as a volunteer and a consultant for countless health agencies, institutes, universities, clinics, and pharmaceutical companies, Departments of Vocational Rehabilitation and the Food and Drug Administration. He was also a fellow of the American Statistical Association.

Clint Miller was one of the panel of original organizers of the Regional Statistical Summer Season (the precursor to the Summer Research Conference) and was a long-time member of SRCOS, and served as president of SRCOS from 1981–1983. In his honor, the first annual Clint Miller Award for Outstanding Student Presentation was given at the 2004 Summer Research Conference in Blacksburg, Virginia, to Amy Bardeen and Thomas McCoy of Clemson University for their poster presentation of "Effects of Endometrial Growth Patterns and Embryo Transfer Time of Pregnancy Rates in an Assisted Reproduction Technology Program."

Michael H. Kutner, PhD



Dr. Kutner received his master's degree in statistics from Virginia Polytechnic Institute and State University in 1962. He taught mathematics and statistics at the College of William and Mary from 1962 through 1967. In 1967, Dr. Kutner received a National Science Foundation Scholarship to pursue his doctoral training at Texas A&M University. After obtaining his doctoral degree in 1971, he took an Assistant Professorship tenure-track position in the Department of Statistics and Biometry at Emory University. Dr. Kutner was appointed Chair of the Department of Biostatistics when Emory formed the School of Public Health in 1990. He was appointed as inaugural Associate Dean for Academic Affairs in 1990 as well. In 1994, Dr. Kutner was recruited by the Cleveland Clinic Foundation as the Chair of the Department of Biostatistics and Epidemiology. Dr. Kutner returned to Emory University in 2000 as Professor of Biostatistics. In 2002, Dr. Kutner became Interim Chair of the Department of Biostatistics in the Emory Rollins School of Public Health. In 2004, he was named Rollins Professor and Chair of the Department of Biostatistics.

He has co-authored two widely-used textbooks and has published over 170 peer-reviewed manuscripts primarily in top-tiered medical journals. His statistical research interests have been notably in the areas of linear models and clinical trials. Dr. Kutner has received numerous awards and honors. He received the Thomas F. Sellers, Jr. Award in 2008 for serving as an excellent role model and mentor from the School of Public Health. In 2011, he received the Dr. Charles R. Hatcher, Jr. Award from the Health Sciences Center for his lifetime of work exemplifying excellence in public health. Dr. Kutner was named a Fellow of the American Statistical Association in 1984, received the ASA Founders Award in 1996 and the W. J. Dixon Award for Excellence in Statistical Consulting in 2011. He received the Mu Sigma Rho for lifetime devotion to statistical teaching award in 2009. In 2002, Dr. Kutner received the Paul Minton Award for service to the profession from the Southern Regional Council on Statistics (SRCOS) and he served as SRCOS President from 2008–2009. Texas A&M's Department of Statistics honored Dr. Kutner in 1984 when he was awarded the H. O. Hartley Award for distinguished service to the profession. In 1997, Texas A&M inducted Dr. Kutner as one of its inaugural recipients into the "Academy of Distinguished Graduates." In 2013, SRCOS named the Junior Faculty Poster Session at the Summer Research Conference in his honor. Junior and isolated faculty have been eligible to apply for Kutner travel awards if they present a poster.

Paul Dixon Minton, PhD (Aug. 4, 1918 – July 10, 2007)

Paul Minton was born in Dallas, TX and grew up in a family with three brothers during the Great Depression. He attended Southern Methodist University (SMU) on an "emergency scholarship" made available to promising candidates from Dallas who otherwise could not afford college. He earned his bachelor's and master's degrees from SMU, Department of Mathematics, with interruption to serve as a cryptanalyst for the FBI during World War II.

Minton continued his studies at North Carolina State University, earning his PhD in 1957 under the direction of Gertrude Cox. While there, he was exposed to R. A. Fisher, H. Hoetelling, W. Cochran, J. Wolfowitz, H. E. Robbins, R. L. Anderson, R. C. Bose, and many other well known statisticians.

Dr. Minton then returned to SMU Mathematics and built a program of statistics courses for students from a wide range of majors. These evolved into the formation of the Department of Statistics in 1961, with Dr. Minton as the founding chair. The PhD degree was instituted in 1966. During this time, Dr. Minton was also named the director of the first computer center at SMU, which housed the Univac 1103, one of the few large scientific computers then available. After chairing the new department for 10 years, Dr. Minton joined Virginia Commonwealth University in 1972 as Dean of the School of Humanities and Social Sciences, a position he held until 1978. In 1979, he formed the Institute of Statistics at VCU and served as its Director until 1987, before retiring in 1988.

Dr. Minton was an active leader in the profession and received numerous awards. He was named a Fellow of the American Statistical Association in 1968 and received the prestigious ASA Founders Award in 1991. He founded the North Texas Chapter of the American Statistical Association (ASA) and was also very active in the Virginia Chapter. Dr. Minton represented SMU on the Southern Regional Education Board Committee on Statistics (precursor of SRCOS) from 1963 to 1972, and chaired this committee 1968–1969. He also represented VCU on this committee from 1979 to 1987.

Dr. Minton was known for his dedication to service in all his roles. In the words of J. M. Davenport, Minton "...was a great champion of those who were in need of assistance, opportunity, and needed encouragement. He had this innate ability to meet you where you were, offer words of wisdom, give you the resources you needed, and push you in the right direction. He was truly a great teacher." In further recognition of his contributions, the SRCOS established the Paul D. Minton Service Award in 1992 to recognize outstanding service to the statistics profession.

References for all biographies

- "Dr. Clint Miller, former Biometry chair, dies," *The Catalyst Online*, Medical University of South Carolina Office of Public Relations, Feb., 2004.
- *The Spectrum*, Virginia Polytechnic Institute, 20(32), June 4, 1998.
- J. C. Arnold, K. Hinklemann, G. G. Vining, and E. P. Smith, "Virginia Tech Department of Statistics," in *Strength in Numbers: The Rising of Academic Statistics Departments in the U.S.*, A. Agresti and X-L. Meng (eds.), Springer-Verlag, 2013.
- Anderson, R.L. (1982). "My Experience as a Statistician: From the Farm to the University," in *The Making of Statisticians*, J. Gani (ed.), Springer, 129–148.
- Taylor, R.L. and Padgett, W.J. (2006). "The Summer Research Conferences and the SRCOS: A Historical Perspective," *J. Statistical Computation and Simulation*, 76(5):373–383, May, 2006.
- "Richard L. Anderson" in *Statisticians in History*, published on-line by the American Statistical Association, (ASA biography url). See `http://www.amstat.org/about/statisticiansinhistory`.
- Brock, D.B. (2008). "Building a Department," *Amstat News*, Sept., issue 375.
- Davenport, J.M. (2008). "Paul Dixon Minton: LSD for Statisticians from a Southern Gentleman", paper from Memorial Session, Joint Statistical Meetings.

The Mission of SRCOS

The mission of SRCOS is to promote the improvement of postsecondary education in statistical science, assist in the development of high quality statistics instruction in elementary and high schools, and promulgate educational activities that improve the quality of statistical practices. SRCOS fulfills this mission by fostering and facilitating cooperation among institutions in its membership region concerned with statistics education.

Specific examples of SRCOS activities in fulfilling its mission are providing forums for communication on effective approaches to solving common problems with statistical training, sponsoring (joint with ASA) annual summer research conferences, and maintaining an electronic network for sharing statistical information among its members.

Acknowledgments

We extend our warmest thanks to all participants and to the following organizers:

2026 SRCOS Officers

Dr. Katherine Thompson	SRCOS President, University of Kentucky
Dr. Whitney Huang	SRCOS President-Elect, Clemson University
Dr. Madhuri Mulekar	SRCOS Past-President, University of South Alabama
Dr. John Wierman	SRCOS Treasurer, Johns Hopkins University
Dr. Norou Diawara	SRCOS Secretary, Old Dominion University

SRCOS Website Administration

Dr. Zhaoxue Tong, Florida State University
Dr. Sy Han Chiou, Southern Methodist University

Scientific Program

Dr. Norou Diawara, Old Dominion University, Chair
Dr. Jeremy Gaskins, University of Louisville
Dr. Zhaoxue Tong, Florida State University
Dr. Sy Han Chiou, Southern Methodist University

Local Arrangements

Dr. Achraf Cohen, University of West Florida, Chair
Dr. Ranadeep Daw, University of West Florida
Dr. Subhash Bagui, University of West Florida
Dr. Jia Liu, University of West Florida
Dr. Shusen Pu, University of West Florida

Co-hosts of the SRCOS 61st Summer Research Conference

Hosted by the Department of Mathematics and Statistics
University of West Florida