Refining a Two-Stage Pipeline for Gene-by-Environment Interaction Discovery: Random-Forest Screening Followed by SOIL-Based Prioritization

Kakon Datta, University of Kentucky

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.

Mapping Inequality: Women Statisticians and the Settlement House Movement

Tisha Prasad, STEM4Change

This talk will examine the connection between the Settlement House movement and the development of social statistics, focusing on the women who used data collection and visualization to drive labor and welfare reform in late nineteenth-century Chicago. At Hull House, reformer Florence Kelley and her colleagues (including Jane Addams, Julia Lathrop, and Agnes Sinclair Holbrook) conducted door-to-door surveys of one of the city's most overcrowded immigrant districts, mapping wages and nationalities block by block. The resulting 1895 publication, Hull-House Maps and Papers, became a landmark in statistical graphics and a model for the modern social survey. Far from a purely academic exercise, this data collection was deployed directly as evidence: Kelley's findings on child labor and sweatshop conditions were presented to Congress and the Illinois legislature, helping secure passage of the state's first factory inspection laws. This talk traces how Hull House women built one of the earliest bridges between statistical methodology and social policy, and considers their legacy for the role of data in advocacy today.

Precision by Design: A Transferable Framework for Sampling Optimisation in Hierarchical Capture-Recapture

Su Na Chin, Universiti Malaysia Sabah

Capture–recapture methods are widely used to estimate population size when complete enumeration is impractical. Yet sampling effort across capture occasions is often allocated by convention rather than statistical efficiency. This study develops a hierarchical Schnabel census in which each primary occasion contains several sub-occasions, such as trapping days or repeated deployments. Under a fixed total effort, the aim is to determine the allocation that minimises the variance of the population estimate. When per-sub-occasion capture probability is constant, convexity arguments and Jensen's inequality show that equal allocation is optimal. Simulations confirm that it generally yields lower empirical and asymptotic variance than skewed or random allocations. The same result holds when capture probabilities follow a common symmetric distribution. However, when capture probabilities decline over time, equal allocation is no longer optimal. Monte Carlo optimisation indicates that more effort should be assigned to later occasions to offset reduced detectability and improve precision. The findings show that efficient design depends not only on total effort, but also on how detectability changes over time. The framework provides practical guidance for designing more efficient capture–recapture studies across ecology, fisheries, conservation, and related fields.

How Biostatistics Research Experience Enriches the Pre-Medical Journey

Kevin Zou, University of Central Florida

Pursuing a career in medicine, public health, or another discipline often requires students to simultaneously build a competitive graduate school application while developing the knowledge, skills, and attributes needed to become future clinicians, scientists, or other professionals. At first sight, biostatistics research may appear distant or unrelated to a future career, but many of the skills it builds are directly transferable. Participation in biostatistics research offers a unique opportunity to cultivate technical, analytical, and writing skills. As an undergraduate student, I will reflect on my experience of working within a biostatistics research team on a project investigating disparities in sleep duration among Florida college students. In this talk, I will share my perspective on how to find the right biostatistics lab and how my research experience – from learning to code with SAS® 9.4 to presenting the work at a conference – supports my pre-medical development and goes beyond the resume. I will conclude with an overview of the broader value of biostatistics research for pre-medical and non-statistics students in general, emphasizing that research extends well beyond merely strengthening applications.