Observation-Weight-Based Strategies for Robust Stochastic Gradient Boosting in Genomic Prediction

Beatriz Gil Comparado, FCTNOVA Math

The presence of data contamination, such as errors or outliers, can severely compromise statistical models, including ensemble-based machine learning methods such as Stochastic Gradient Boosting (SGB), particularly in high-dimensional genomic prediction and selection studies. Although contamination may occur at both the response and covariate levels, this work focuses on response contamination and evaluates the robustness of the classical SGB method via simulation. Using a synthetic animal breeding dataset from the literature, we consider several plausible contamination scenarios and investigate robust adaptations of SGB based on observation-weighting strategies designed to enhance model resilience. Our study highlights the impact of data contamination on genomic prediction and explores robust alternatives to classical SGB that are suitable for routine use in breeding applications.

A journey of pivoting: From Local Research Experiences to Global Data Science Connections

Cherlynn Dumbura, CeSHHAR

Background:

Early career researchers are increasingly building careers across disciplines, sectors, and global research networks. For many, the challenge is not lack of skill but knowing how to translate existing training into a new professional direction. This talk designed for ECRs who are considering a pivot into data science, climate and health research, research operations, or scientific leadership, and who want to reposition themselves with clarity and confidence.

Body:

Using my journey as a Zimbabwean data scientist and public health researcher, I reflect on my journey from Biochemistry to Biostatistics to Data Science and beyond. I reflect on the process of building a career at the intersection of statistics, maternal and newborn health, climate change and research coordination in a cross sectoral consortium.

I reflect on how transferable skills from each point in my journey including leadership roles in scientific communities have shaped me in terms of strategy and scientific communication.

The talk offers a practical framework for career transition: Identifying an intellectual anchor, auditing transferable skills, selecting a technical edge, building visible proof of work, and using professional networks to access new opportunities. I also reflect on my own next phase of growth, including the integration of geospatial analysis, artificial intelligence, and climate and health decision-support tools into my research portfolio.

Teaching Statistics in the Age of AI: Beyond Calculation in Health Sciences Education

Camila Martins, Federal University of São Paulo

Statistics plays a central role in scientific knowledge production and evidence-based decision-making. In health sciences education, statistical literacy is essential for interpreting research findings, critically appraising the literature, and supporting clinical decisions.

Despite its relevance, teaching Statistics to health sciences students remains challenging. Many students experience anxiety toward discipline and struggle to connect statistical concepts with professional practice, often reducing learning to the memorization of formulas and statistical tests rather than meaningful reasoning.

The expansion of generative AI has added new complexity to this scenario. Statistical analyses, code generation, and result interpretation can now be partially automated, raising an important question: what should remain at the core of Statistics education?

This presentation discusses Statistics education in undergraduate health sciences programs considering these transformations. It argues that teaching should move beyond procedural execution and prioritize statistical reasoning, critical thinking, and data literacy. Special emphasis is placed on competencies such as formulating relevant questions, assessing data quality, recognizing methodological limitations, and critically interpreting evidence under uncertainty.

In the age of AI, Statistics education should focus less on calculations and more on preparing students to think critically about data, uncertainty, and evidence.

Observation-Weight-Based Strategies for Robust Stochastic Gradient Boosting in Genomic Prediction

Beatriz Gil Comparado, FCTNOVA Math

The presence of data contamination, such as errors or outliers, can severely compromise statistical models, including ensemble-based machine learning methods such as Stochastic Gradient Boosting (SGB), particularly in high-dimensional genomic prediction and selection studies. Although contamination may occur at both the response and covariate levels, this work focuses on response contamination and evaluates the robustness of the classical SGB method via simulation. Using a synthetic animal breeding dataset from the literature, we consider several plausible contamination scenarios and investigate robust adaptations of SGB based on observation-weighting strategies designed to enhance model resilience. Our study highlights the impact of data contamination on genomic prediction and explores robust alternatives to classical SGB that are suitable for routine use in breeding applications.

A journey of pivoting: From Local Research Experiences to Global Data Science Connections

Cherlynn Dumbura, CeSHHAR

Background:

Early career researchers are increasingly building careers across disciplines, sectors, and global research networks. For many, the challenge is not lack of skill but knowing how to translate existing training into a new professional direction. This talk designed for ECRs who are considering a pivot into data science, climate and health research, research operations, or scientific leadership, and who want to reposition themselves with clarity and confidence.

Body:

Using my journey as a Zimbabwean data scientist and public health researcher, I reflect on my journey from Biochemistry to Biostatistics to Data Science and beyond. I reflect on the process of building a career at the intersection of statistics, maternal and newborn health, climate change and research coordination in a cross sectoral consortium.

I reflect on how transferable skills from each point in my journey including leadership roles in scientific communities have shaped me in terms of strategy and scientific communication.

The talk offers a practical framework for career transition: Identifying an intellectual anchor, auditing transferable skills, selecting a technical edge, building visible proof of work, and using professional networks to access new opportunities. I also reflect on my own next phase of growth, including the integration of geospatial analysis, artificial intelligence, and climate and health decision-support tools into my research portfolio.

Teaching Statistics in the Age of AI: Beyond Calculation in Health Sciences Education

Camila Martins, Federal University of São Paulo

Statistics plays a central role in scientific knowledge production and evidence-based decision-making. In health sciences education, statistical literacy is essential for interpreting research findings, critically appraising the literature, and supporting clinical decisions.

Despite its relevance, teaching Statistics to health sciences students remains challenging. Many students experience anxiety toward discipline and struggle to connect statistical concepts with professional practice, often reducing learning to the memorization of formulas and statistical tests rather than meaningful reasoning.

The expansion of generative AI has added new complexity to this scenario. Statistical analyses, code generation, and result interpretation can now be partially automated, raising an important question: what should remain at the core of Statistics education?

This presentation discusses Statistics education in undergraduate health sciences programs considering these transformations. It argues that teaching should move beyond procedural execution and prioritize statistical reasoning, critical thinking, and data literacy. Special emphasis is placed on competencies such as formulating relevant questions, assessing data quality, recognizing methodological limitations, and critically interpreting evidence under uncertainty.

In the age of AI, Statistics education should focus less on calculations and more on preparing students to think critically about data, uncertainty, and evidence.