IBC 2024 Atlanta, Statistics In Practice Showcase

Organizer: Rafael de Andrade Moral, Maynooth University

Georg Heinze, Center for Medical Data Science; Institute of Clinical Biometrics, Medical University of Vienna
Theresa Ullmann, Center for Medical Data Science; Institute of Clinical Biometrics, Medical University of Vienna

Abstract

This "Statistics in Practice" session is presented on behalf of the STRATOS initiative's Topic Group 2, which focuses on selection of variables and functional forms in multivariable model building.

In many scientific fields, statistical models play a crucial role in describing, predicting, and explaining outcomes with empirical data. Two key and interconnected challenges in model building are selecting the appropriate variables and determining the functional forms of continuous variables within the model.

The first session will delve into these challenges, with a strong focus on aligning the model-building process with the intended purpose of the model. We will emphasize that depending on the model's objective, the use of variable selection algorithms may either be helpful or ill-advised. Various variable selection methods will be reviewed, highlighting how their application can affect the stability of a model and introduce additional, often overlooked, uncertainty. Additionally, we will explore techniques for handling non-linear functional forms of continuous variables. Although variable and functional form selection is a frequent issue in model building, few algorithms address this combined task, and there is limited knowledge about their relative performance. Some of the existing proposals will be presented and discussed.

In the second session, we will take a deeper dive into the topic of variable selection, using a simulation study to empirically examine the properties of several popular algorithms across different scenarios. The results of this study will be presented via an interactive application, allowing for a visual comparison of methods and an overview of their performance. This dynamic presentation will help attendees grasp the nuances of variable selection algorithms. Finally, we will conclude with evidence-based recommendations for practicing statisticians, informed by our simulation results, on the effective and safe use of variable selection methods in real-world applications.