

Final Project Proposal

Marwin Carmo

Background

Variability is important and conveys information not provided by average achievement alone. Carmo et al (2024) propose the Spike-and-Slab Mixed-Effects Location Scale model (SS-MELSM) methodology to identify clustering units (students, classrooms, etc.) that exhibit unusual levels of residual variability—such as consistency or inconsistency—in academic achievement. This method is implemented in the R package `ivd`. The main information provided by this package are the Posterior Inclusion Probabilities (PIP) which quantifies the probability that a given random effect is included in the residual variance (scale) model, conditional on the observed data. A high PIP (i.e., $> .75$) indicates strong evidence that the random effect is necessary to explain the data. Evidence for retaining the random effect is evidence of unusual variability. After identifying the (in-)consistent schools researchers might want to investigate this clustering units to understand what distinct these clusters from the others. For example, how does their variation compare to the rest? How are the students within the schools performing on average? Visualization can help quickly answer some of these questions and provide initial basis for further investigations (i.e., which schools are they interested in conducting further investigation or intervention).

Research question

The broad research question motivating the SS-MELSM is how to identify and isolate clusters (such as students, classrooms, schools, etc.) that display unusual amounts of residual variability. Given the purpose of this class, I will define the research question as how to visualize these clusters such that it is clear from the visualization what makes these units unique in relation to the others. The final goal is to enhance the type of visualization currently provided by `ivd`.

Method

In this project I will work with the example data included in `ivd`. It consists of standardized math scores from 11,386 11th and 12th-grade students across 160 schools. The data are the

posterior estimates of the scale model, which models the scale as function of fixed effects of student and school-level SES and a random intercept. Specifically, I will use the PIPs, the estimated random effects standard deviations and the estimated math scores.

Visualization plan

Challenges and barriers