Final Project Proposal

Marwin Carmo

Background

Variability in academic achievement provides crucial information beyond average performance alone. The Spike-and-Slab Mixed-Effects Location Scale model (SS-MELSM) methodology identifies 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 package's main output is the Posterior Inclusion Probability (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, suggesting unusual variability in that cluster.

Once (in)consistent schools are identified, researchers may want to investigate these clusters to understand what distinguishes them from others. For example, how does their variation compare to the rest? How are students within these schools performing on average? Visualizations can help answer these questions quickly and provide a foundation for further investigation.

Research question

The broad research question motivating the SS-MELSM is: *How can we identify and isolate clusters that display unusual amounts of residual variability?* For this project, I will focus on a more specific question: *How can we visualize these clusters to clearly highlight what makes them unique compared to others?* The final goal is to enhance the visualizations currently provided by the `ivd` package.

Method

For this project, I will use the example dataset included in the `ivd` package, which consists of standardized math scores from 11th and 12th-grade students across 160 schools. I will work with posterior estimates from the scale model, which models the scale as a function of fixed effects (student- and school-level socioeconomic status) and a random intercept. Specifically,

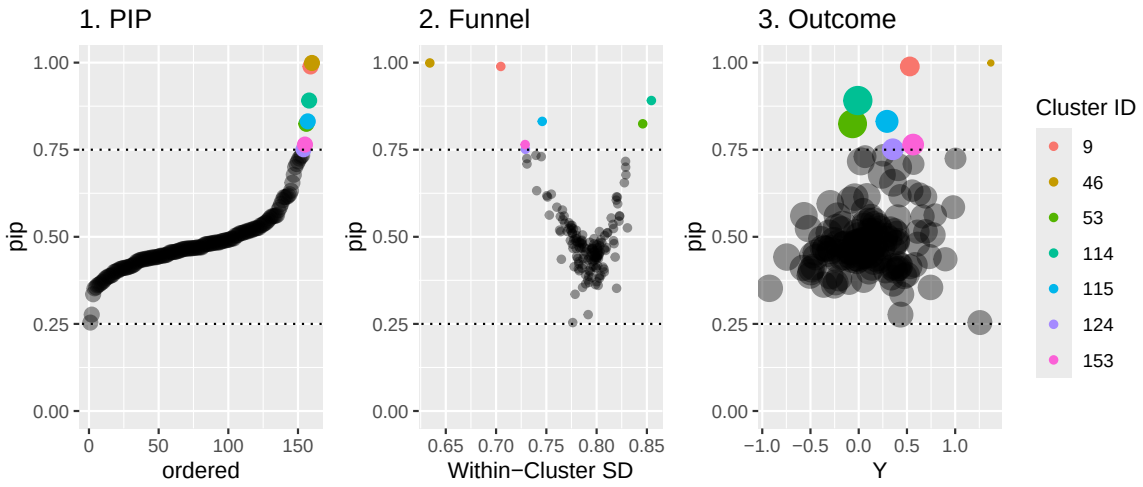
I will use the PIPs, the estimated random effects standard deviations and the estimated math scores.

Visualization plan

The package currently provides three types of visualizations, all of which display PIPs for a given random effect on the vertical axis:

1. `pip`: Displays the sorted school index on the horizontal axis.
2. `funnel`: Plots PIPs against the estimated within-cluster standard deviation.
3. `outcome`: Shows the observed outcome value on the horizontal axis, with point size determined by the estimated within-cluster standard deviation.

While these plots are simple and allow for user customization, there is room for improvement. For example, providing a better description for the axes, using a better color palette, implementing a smarter label scheme, or developing entirely new visualization formats.



Challenges and barriers

The most significant challenge in this project is labeling the clusters effectively. When a small number of high-PIP clusters are identified, annotating each point with the cluster ID works well, as demonstrated in the SS-MELSM pre-print. However, as the number of clusters exceeding the PIP threshold increases, the points and labels tend to clump together, making the visualization cluttered and difficult to interpret. Additionally, in the `outcome` plot, the size of the points can become hard to distinguish when the within-cluster standard deviations are similar.