

Written Proposal and Modeling Approach

Shannon Jay McElhenney, Raymond Tana, Prabhat Devkota, Shrabana Hazra, Jung-Tsung Li

1 November 2024

Modeling Problem

We'd like to analyze the effect of *proximity to universities* on *AP Test Outcomes*. It is well known that there is a strong relationship between AP outcomes and average income. Proximity to quality institutions with high-school outreach, educational, and/or dual enrollment programs may have a secondary effect on performance on AP exams.

The data currently available through AP only provides performance data at a state-level, which is likely too coarse for us to notice proximity-based effects from higher institutions. Still, we will validate some of our hypothesized relationships on this state-level data.

For a few states, we were able to find comprehensive data from over the past five years at a school district- or county-level. We selected Georgia, Massachusetts, and Wisconsin individually for their comprehensive data, and for forming a diverse triad of states. Within each state, we will perform a similar analysis to the national level, collecting all data from 2019, 2020, 2021, 2022, and 2023.

Relevant Features

We plan to model outcomes on AP exams (i.e., the *passing rate*, which means the proportion of students who scored 3, 4, or 5 on at least one of their exams) as a function of a few key features.

- *Average per-capita income* within a county
- Measure of *university presence* aggregated into several university classifications (R1/R2; undergraduate; or professional/associate), affiliations (public; private), size, and/or selectivity
 - Simpler method: tally of universities also in that county
 - More complex method: tallies of universities in a R -mile radius of that county
- Hypothetically "irrelevant" features
 - *Number of dorm beds*
 - *Semantic content of school district name*