Title: Impact of Internet Access and Pandemic Learning Models on Educational Outcomes in Pennsylvania

Authors: Mena Tetali, Sara Murley, Samantha Rudravajhala

Problem Statement:

How did the amount of time districts spent using virtual or hybrid learning during COVID affect post-pandemic test scores in areas with different levels of access to the internet?

We intend to look specifically at the state of Pennsylvania using county-level data. We will look at broadband coverage by county, as well as the average amount of time that school districts within the county spent doing virtual learning during the pandemic. We will compare test scores (separately math and reading) from before and after the pandemic to see how educational outcomes in different counties were impacted based on the learning model and internet access.

Background:

In 2020 the Covid-19 pandemic swept the country. During this unprecedented time, schools around the US were forced to change course and adapt to new learning methods. The most common way of learning during the pandemic was through remote or online learning. The shift in methods was tough not only for teachers but for students as well as many had trouble focusing and were less likely to ask for help if they were struggling. This shift also exposed the disparities in internet access between rural and urban communities. In the years following the pandemic, studies have shown that test scores have decreased.

On March 19, 2020, Tom Wolf, the governor of Pennsylvania, called for all "non-life sustaining businesses operations and services" to shut down. This included schools. Pennsylvania is a very diverse state with both urban and rural areas. Given the difference in broadband access across the state, the effect of online learning was different in each county and even in each district. These effects would eventually be seen in the following school year.

Data:

We will use a combination of various datasets for our analysis. First, we intend to scrape this interactive map to get broadband data by county or find the CSV data behind the graphic. Alternatively, this dataset has a value for the number of households in the county that have access to the internet.

Second, we will use test scores from before and after the pandemic to assess changes in educational outcomes. Pennsylvania administers the Pennsylvania System of School Assessment (PSSA) annually in grades 3-8. We will use total percentages of students across all test-takers that fall into each category of performance to measure the educational outcomes of each district, and then calculate the averages across the county using this dataset.

Finally, in order to compare different COVID-19 policies, we will use data on how long districts were relying on virtual learning. This dataset has data for each month for each school district in the state. We will count the number of months that each district was virtual and then get averages across counties.

Methodology:

The majority of our project will involve data wrangling in Pandas. We will need to clean our datasets, aggregate data by county, and join the datasets together to create a final dataset. In our final dataset, we will have one row per county with an indicator of internet access, average test scores pre- and post-pandemic (for math and reading), and the average number of months schools in that county spent using virtual or hybrid learning. We will potentially use scraping techniques to generate our initial dataset by obtaining data from the interactive state broadband map.

After we obtain our final dataset, we can conduct exploratory data analysis on our data. We will use data visualization packages like plotnine, ggplot, and seaborn to create maps of the state and charts of student performance. After visualizing the data, we will conduct statistical analysis using linear regression to see the relationship between time spent virtual, internet access and test scores.

In order to identify causal relationships between COVID-19 policy and test scores, we can use other statistical methods such as matching across districts that share certain characteristics (especially internet access) or using spatial matching using neighboring districts.

Success Statement:

Our project aims to examine how internet access interacts with COVID-19 return-to-school policies to shape educational outcomes. Guided by our shared commitment to health and education policy, we anticipate finding that limited broadband access has a pronounced negative impact on test scores, particularly in districts that delayed returning to in-person instruction. Insufficient internet access likely amplifies the adverse effects of stricter COVID policies on student performance.

Through this analysis, we hope to urge future policymakers to prioritize broadband internet access as a fundamental factor in pandemic preparedness. Ensuring robust internet access could allow districts to uphold public health measures without compromising educational continuity. Additionally, further research should consider examining the data across different demographics, focusing on rural and lower-income areas where connectivity and poverty challenges may intersect. Analyzing racial demographics will also be important to see if certain groups face more significant broadband access limitations. By quantifying these issues, we aim to validate our hypothesis and support future efforts to expand broadband access within underserved communities.