MA346 Final Project Report – Urbanization Data – Debayan Sen

In this project, I looked at data from the UNICEF website concerning urbanization. There are two sheets in this dataset. The first sheet has urbanization data that looks at the total, urban, and rural sections of each country. The countries in this dataset are from Asia, Africa, Latin America and the Caribbean, and Europe. There are different indicator categories that are evaluated for each country in their rural, urban, and total sections. Examples of these categories include "Birth Registration Rate" or "Primary Education Completion Rate". The second sheet is similar, except there is no "total" residence category, and instead there are five wealth quintiles (Q1,Q2,Q3,Q4,Q5) that the data looks at. Each country appears ten times in this dataset, five times to represent each quintile in the urban category and five times to represent each quintile in the rural category. Indicator scores are provided on both sheets, and these are a measure of how well a country does in respect to their respective categories. I looked at these sheets sequentially to make meaningful conclusions about the data. This link will bring you to the code I used to analyze the data along with explanations, and this link will bring you to an interactive web app I made to visualize the data.

The first sheet – Urban versus Rural

I first looked at the difference between urban and rural indicator scores. The first graph in the interactive web app above allows the user to select any amount of indicator categories which will display both the average urban indicator score and rural urban indicator score as a grouped bar graph. Here we see that there does seem to be an advantage that urban areas have over rural areas. To see which indicators where urban areas were better than rural areas, I ran a t-test on each category and interpreted the p-value at 5% significance. A t-test is a statistical test that tells us whether we can conclude two population means are truly different or not. From the t-tests, I saw that across every category there did seem to be a significant difference except for DPT3 coverage (which refers to the three doses of diphtheria, tetanus toxoid and pertussis vaccine). I also looked at the difference in indicator scores across the categories in just urban areas, which revealed how different the indicator values varied between indicators.

Second sheet – Wealth Quintiles

The second sheet had each country appear ten times as there were both rural and urban areas, both each with five separate quintiles. I wanted to see the difference between urban and rural areas again, but this time across wealth quintiles. The second graph in the interactive web app allows us to do this, where the user enters a single indicator category and then the graph displays each of the quintiles and the average urban and rural indicator value. Here, we see a similar trend that we saw in the first sheet where urban areas tend to have a higher indicator score than rural areas (important to note that for the Under 5 Mortality rate, a higher indicator value shows worse conditions). Once again, I conducted a t-test on the combination of all combinations of indicator categories and quintiles. This resulted in 45 t-tests, where the only time I could conclude there was no difference between the average indicator score was for mainly DPT coverage.

Takeaways

The main takeaway from both sheets of data is that there does seem to be an "Urban advantage" that exists across the countries in the dataset except for in the DPT3 coverage area. This means that DPT3 vaccination efforts are doing well in reaching rural areas, however, there needs to be an increased focus in providing more resources to rural areas to increase access to better quality education, water sources, and health services.

Areas for more data exploration

One area I did not explore in this data is the difference between different continents such as Europe, Asia, and Africa. It would be beneficial to see the differences in urban and rural indicator values in these distinct locations and in relation to one another to see how a location's specific continent also affects their indicator scores. Another area of exploration is looking at the difference between wealth quintiles and focusing on how to resolve disparities that exist between these quintiles.