Group 28 – 1st Project Check-in Discussion

Mohammad Haddadnia

PRAISE

- I really like the question you have selected. It is really practical and is a very important public health related problem.
- The dataset you have selected seems very excessive and other variables that are present should give plenty for room to adjust for confounding and other sorts of interactions.
- I also love the fact that your project is already organized on GitHub to keep everything in one place and have an easier time to share things with each other (and the teaching team)

ADVICE

- In the dataset, it seems like some countries are missing data for only some of the variables, so I suggest you choose a strategy that minimizes the number of countries/datapoints you end up dropping. Perhaps you can first decide which variables you would want to include and drop only the data points that have missing values in the remaining variables.
- The dataset seems to separate female and male under 5 mortality. Perhaps you can look into whether there is effect modification (or some sort of interaction) of sex in your question. If you have data on females vs. males, it might not be a bad idea to take advantage of it and perform some more modeling, before collapsing it into a single under 5 mortality variable.
- You mention " higher income" in your hypothesis, which I assume suggests you are considering income in a categorial manner? If so, make sure you mention that somewhere and how you categorized the variable. My advice is to be more specific in this regard: do you have a threshold for what is considered high? E.g. "higher income (>200K USD) is associated with lower mortality rates."
    - Also, another variable in your dataset in unemployment, which might be related to income. So perhaps you can perform some analysis here to test whether you want to include both, or if not, which one would you pick.
    - Also, there are two GDP per capita columns in your dataset. Be more specific on which one you are choosing and why.
- For stratification, I like that you want to look at different regions and groups of countries, but again, be specific with how you are defining these regions. My advice would be to perform some literature search to see what type of grouping makes the most sense for YOUR question.

CONCERNS

- The data seems to be from 2019 only (at least the one you've uploaded on GitHub), so how generalized can your model be to other time points? Is your question going to be specific for 2019 only?
- Since the data is from 2019 only, it might be possible that there is confounding by some geopolitical events that are not present in your datasets (e.g. wars, political

tension/sanctions, etc.). Perhaps you can also look into the political state of the countries that are present in your dataset in 2019 (or certain periods of time if you go beyond just 2019).

If you have any question or need clarification on anything, feel free to reach out. I'd be happy to chat!