Antonio Ruiz Fernandez

IE Project Individual Reflection

I am glad that we selected the COVID-19 dataset for this project, there were a lot of potential important questions we could tackle with regression models, it would have been great to have had more time to analyze different response-covariate combinations. As a summary, we decided to tackle two, technically 3, questions. What is the effect of a country's HDI on the country's death by COVID-19 percentage? What is the effect of a country's vaccination percentage on the country's death percentage? How does this compare to a country's fully vaccinated population percentage and its effect on the country's percentage of population that died from COVID-19?

The key highlights from our conclusion were that we found a positive linear relation between the percentage of the population in a country that died from COVID-19 and that country's HDI, we got this insight from our t-tests after determining that the model had minimal regression violations. For the other two models, the aggregated versions showed better results after we did some transformations, allowing us to have minimal violations to our regression assumptions. Therefore, we can conclude with certainty that there is a small positive relationship between the percentage of deaths due to COVID-19 in a country and that country's squared percentage of vaccinated people, and a larger positive relationship between the same response and that country's squared percentage of fully vaccinated people.

These are very important data-based conclusions that are mathematically robust. Even if validating our regression assumptions can be sometimes subjective our transformed models showed almost no deviations from the assumptions, resulting in very valuable conclusions.

The workload was most of the times evenly spread out within the team. I mostly worked on the first model (percentage of deaths with respect to HDI) in most of the IE submissions, as in the introduction for the first IE and the final introduction segments for the final IE. For this first model, I had to do some data transformations to aggregate the data based on country to get cleaner results. After this I did the regression and interpreted the results of our tests. I also tested the validity of our regression assumptions for this model. Team communication was sometimes slow with Aadam because he took a while to respond, but aside from this it was smooth. Aditya did a lot of work for the other questions on vaccination percentages, this was a good amount of work because we had to deal with transformations for these models. Christopher did most of the conclusion and test analysis in the project, he also worked on the second and third regression models in the final IE. Aadam was not able to work in one of the IEs and did not complete the part he promised for another IE, however for that IE he did all the transformations necessary to the dataset to get the covid_new.csv, so he still did (I would say) a good amount of work. Aadam worked on all IEs after that without a problem and did most of the conclusion work for the final IE.