Regression Proposal

FURTHER DATA PROCESSING:

From my previous graph on the world covid trend, and based on our current knowledge about the covid trend. We know the data covers covid data from inception roughly till its peak worldwide. Thus, for each country, we select three time points to represent: beginning phase, developing phase, outbreak phase. Discard data with missing value at these time points. Create a dummy factor variable to indicate the three phases.

Then, we create another variable: mortality rate= total death/ total case. I think we should regress against total case and mortality, instead of total case and total death. Since the absolute numeric value is not informative. Countries like China may have large total death due to large population, but the mortality rate is low.

Next, since during the data processing step, I noticed that population effect on HDI is different for developing and developed country , we create another factor variable to indicate whether the country is developing, developed or underdeveloped. We can call the variable develop

Regression:

Model1 is like GDP per capital~ total case + mortality + population + HDI + STI+ STI: Phase, we want to add the additional interaction term because we want to learn the effect of STI in different phases. For instance, we may want to know whether an extremely strict STI at the beginning of covid is beneficial to the economy. Or whether strict STI will negatively impact GDp for all phases.

Model2 is like HDI~ total case + total death+ population + STI+ GDP per capita +GDP per capital: population + total death: develop.

We add interaction between gdp per capital and population because we want to consider size effect. Large population with relatively low per capital gdp can still produce very large total gdp and influence HDI. Also, we include interaction total death: develop in order to study whether more number of death is beneficial to poverty alleviation (HDI) .

We can perform both robust regression and ordinary regression like the paper suggested.

If we want to add more interaction terms,( if you have any other effects you want to test ), we can do ridge regression or AIC for variable selection.