With continuously changing abilities and accumulating mutations, SARS-CoV-2, the virus that causes COVID-19, constant evolvements and accumulated mutations in its genetic code over time. The emergence and quick spread of the alpha, beta, and delta SARS-CoV-2 VOCs have generated continuous waves of infection in the past two years. The virus has brought tremendous shocks to the supply side of the economy and resulted in millions of deaths around the globe, representing an unprecedented tragic loss of the whole human society.

By analyzing the Covid-19 Case Surveillance Public Use Data from the Centers for Disease Control and Prevention, our project aims to identify the primary factors that are sensible to the effects of Covid-19. We mainly focus on samples in North Carolina and ignore the individual observations which have missing/unknown live status records. Focusing on a single state would help eliminate potential time-invariant effects among different states.

We hope this project will bring suggestive policy implications by identifying the most vulnerable groups against the Covid-19 virus among the population. Hopefully with our convincing results, the medical facilities would be able to allocate resources, such as hospitalization and medical aids, to the appropriate groups efficiently. Also, the government can assign social welfare benefits and designate priorities for vaccination by understanding which group is most vulnerable to the virus.

1. Regression model: 咋build,为啥能这么build
2. Decision Tree:
3. Double-robust estimator

Q1: What are the primary factors you intend to test initially (what treatments)?

Q2: Policy implications based on your results?

Q3: What methods, such as the linear regression model, do you implement for testing?

Q4: Why choose this model? What are the (dis-)advantages of choosing linear regress/decision tree/…/ over others?

Q5: How do you compare the results from different models? Which are the key coefficients/parameters you look at?

Q6: Why choose North Carolina? Why not choose the United States?

Q7: Why 2020?

Eliminate potential effects from vaccination.

Q8: Any limitations?

Q9: Any past literatures? What are the differences between this project and them?

[1] Paper 1: **Predictors of Death Rate during the COVID-19 Pandemic**

Can check the potential factors affecting the death rate. OLS/Two-stage/Lasso

Link: <https://www.mdpi.com/2227-9032/8/3/339/htm>

[2] Paper 2: **Determinants of COVID-19 Death Rate in Europe: Empirical Analysis**

Three hypothesis, mainly linear regression

[3] Paper 3: **WHAT DOES AND DOES NOT CORRELATE WITH COVID-19 DEATH RATES**

Q10: What is your datasource?

就我们可以用这个tau dr去estimatecasual effect