COVID-19 Mortality Drivers in Brazil: Prioritizing Patients With COVID Symptoms To Reduce Mortality

Hommez, Zara (School: Portola High School)

Problem: The COVID-19 pandemic put hospitals into a state of frenzy attempting to handle the sheer quantity of patients. This project aims to predict mortality likelihood of patients that test positive for COVID-19 solely based on pre-existing conditions, demographics and symptoms, thus providing hospitals the means to identify higher risk patients before running any diagnostics and act accordingly. Solution: A Brazilian database with nearly 1.2 million patient registration forms from the 1st COVID wave was utilized to run machine learning and Cox regression models that predict mortality likelihood. Extensive data cleaning, imputation of missing values and feature selection of the most impactful pre-existing conditions, symptoms and demographics was done to boost accuracy of these models. Results: The top performing XGBoost model achieved an AUC of 83.8% with age by far the top drivers in all models, followed by quality of local healthcare and directly relevant respiratory symptoms. Mortality risk was clearly exacerbated by comorbidities like liver disease or diabetes. Conclusions: My hypothesis that not only comorbidities, but also demographics play a big role in the mortality rate was validated by the models. It also confirms feasibility to build practical applications to prioritize patients for further diagnostics in times of capacity constraints.

Awards Won:
American Statistical Association: In-Kind membership to ASA for all winners, including honorable mentions
American Statistical Association: Second Award of $1,000