Exploring COVID-19 Deathrates and Deathtolls - Handout

Jean-Sebastien Paul

Explanation of the variables in the raw dataset that were used

Unifying Data

Countries and territories *countriesAndTerritories*

Date *Date*

Recovery Data

COVID-19 cumulative recovery numbers *recoveries*[1]

Predictor Data

Population density *Density*[2]

Urban Population *Urban.Pop*

2020 Population *Population.*2020

Median age *Median.Age*

Hospital Beds/ 1000 citizens *Hospital.Bed*[3]

Sex ratio by age (0-14, 15-24, 25-54, 54-64, 65+) $sex14, sex25, sex54, sex64, sex65plus$[4]

Overall population sex ratio *Sex.Ratio*[5]

Overall and for both sexes death rate from lung disease $lung, Female.Lung, Male.Lung$[6]

2018 Sex %Female *Females.*2018[7]

2018 Gross Domestic Product *GDP.*2018[8]

Crime index score *Crime.Index*[9]

Smoking Rate (2016) *Smoking.*2016[10]

[1]This was taken from the John Hopkins University Center for Systems Science and Engineering https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_recovered_global.csv

[2]https://www.worldometers.info/ - Data on Density, Population, Median Age, Urban Population

[3]https://data.worldbank.org/indicator/SH.MED.BEDS.ZS

[4]https://en.wikipedia.org/wiki/List_of_countries_by_sex_ratio

[5]https://data.worldbank.org/indicator/SP.POP.TOTL.FE.ZS

[6]https://www.worldlifeexpectancy.com/cause-of-death/lung-disease/by-country/

[7]https://data.worldbank.org/indicator/SP.POP.TOTL.FE.ZS

[8]https://data.worldbank.org/indicator/NY.GDP.MKTP.CD

[9]https://worldpopulationreview.com/countries/crime-rate-by-country/

[10]https://ourworldindata.org/smoking#prevalence-of-smoking-across-the-world

Testing Data

Cumulative number of COVID-19 tests $Total.tests$[11]

Case and Death Data

COVID-19 new cases $cases$[12]

COVID-19 new deaths $deaths$

COVID-19 total cases $cases2$

COVID-19 total deaths $deaths2$

Major Conclusions Briefly

Binomial and Poisson models are not succesful at modelling the time series data of COVID-19 death toll

The distribution of COVID-19 death rate counts is well approximated by the heavy tail Burr distribution

There is a significant mean difference in death rate between rich and poor countries[13] with poor countries having a lower deathrate. This is perhaps because poorer countries have less testing and are therefore less able to determine who has died from COVID-19, to such a great extent.

Crime and high death rates are not independent.

Death rate is actually not strongly correlated much at all with the other variables. Perhaps this indicates that deathrate is relatively constant. Interestingly, we dont see high degrees of similarity within most of the variables.

We were able to fit a logistic regression model of deathrate onto the interaction terms of all the available numeric variables, adjusted for multicollinearity, to the second power. This had R^2 value of 0.5249956 and adjusted R^2 very close at 0.5060354, indicating that the model did not overfit and strongly explained some of the variance. It should be noted that in diagnostics, a single point's cook's distance was worrying and could have compromised the model.

We were able to fit a forward/backward stepwise linear model (on a linear regression model of deaths onto the interaction terms of all the available numeric variables, adjusted for multicollinearity, to the second power with AIC criterion). This had a strong R^2 of 0.9443243 and similarly strong adjusted R^2 of 0.9424646, indicating that the model did not overfit and strongly explained much of the variance. It should be noted however that diagnostic plots were not promising with the fitted values versus residuals plot not looking random on the left hand side (although it was well spaced as fiited values increased) and the standardized residuals not following a normal distribution. Moreover, an non constant variance test indicated heteroscedasticity. These violations of linear regression assumptions could compromise the model.

[11] https://ourworldindata.org/grapher/full-list-total-tests-for-covid-19

[12] https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide - all case and death data

[13] Where rich is defined as being above mean GDP/capita