

# Homework 1: Estimating Covid-19 Deaths

Neeraj Sharma

4/7/2020

## Assignment

Please submit: the answers to the questions, your code, and your dataset. The code you provide should reproduce your model. All of these should be submitted via canvas.

The Governor of Illinois, J. B. Pritzker, has decided that a key input to public policy is knowing how many people will die from Covid-19 in the near future. He has asked you to estimate the total number of official Covid-19 deaths that will be officially recorded in the state of Illinois by April 21 and by May 31.

To fulfill that request, you will need to assemble a data set, do estimation based on that data, and have some sort of theoretical model in your mind to extrapolate out to the future.

- Describe the data set that you chose to assemble and the rationale behind the choices you made in deciding what data to use.
- Describe the model(s) that you settled on for estimation. What was your logic for using that/those particular models?
- Provide an exact number which is your prediction for cumulative official Illinois Covid deaths through April 21
- Provide an exact number which is your prediction for cumulative official Illinois Covid deaths through May 31
- How did you get from the estimates in (2) to the predictions in (3) and (4)?
- You don't have to provide exact numbers, but discuss what you think the standard errors associated with your estimates might be, and your rationale for thinking those would be the standard errors.
- Make exactly one pretty picture/graph/slide that you would show to the Governor to allow him to easily understand what he should be expecting in terms of Covid deaths.

## Introduction

In December 2019, scientists in China reported the discovery of a novel coronavirus originating from a wild seafood and exotic animal market in the city of Wuhan, Hubei Province, China. Over the subsequent months, the virus spread over the world, infecting individuals on all populated continents and in nearly every country. The assignment given is to provide a prediction of deaths that might occur by April 21 and by May 31 in the state of Illinois for consideration by JB Pritzker.<sup>1</sup>

```
library(tidyverse)
library(readr)
library(curl)
```

## Describe the data set that you chose to assemble and the rationale behind the choices you made in deciding what data to use.

I draw on data provided by the Johns Hopkins Center for Systems Science and Engineering (CSSE). The CSSE provides detailed data on COVID cases in the United States all the way down to the City level. <https://github.com/CSSEGISandData/COVID-19>.

```
# Pulls in the most recent version of COVID 19 deaths from CSSE github repo.
# A static version dated April 8 is saved under "Raw Data" subfolder.
```

---

<sup>1</sup><https://www.nytimes.com/article/coronavirus-timeline.html>

```

covid_url <- "https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_
raw <- read_csv(curl(covid_url))

# Create tidy DF of COVID data for IL only
IL_only <- raw %>%
  filter(Province_State == "Illinois") %>%
  select(-UID, -iso2, -iso3, -code3, -FIPS, -Province_State, -Country_Region, -Lat, -Long_) %>%
  pivot_longer(c(-Admin2, -Combined_Key), names_to = "date", values_to = "cases") %>%
  mutate(date = as.Date(date, "%m/%d/%y"))

```