Lab 2

Srishti Mehra, Andi Morey Peterson, and David Djambazov

11/14/2020

Introduction

A common recommendation from doctors, scientists, and politicians is to wear face masks or other facial coverings to combat the spread of COVID-19. The World Health Organization states: "Masks are a key measure to suppress transmission and save lives. Masks reduce potential exposure risk from an infected person whether they have symptoms or not. People wearing masks are protected from getting infected. Masks also prevent onward transmission when worn by a person who is infected." Because we have not actually performed a controlled study/experiment, we want to build an descriptive model with data available about the current pandemic.

Many states have issues mandates for their citizens to wear masks in public and additionally, many states have issues mandates for employees who interact with the public to wear masks. Now that cases have surged in the United States, we ask: *Do state-wide facemask mandates correlate with the reduction in the amount of deaths in that state?*

Because these mandates and other variables occur in a different times during the nine-month period, we will operationalize the variable "deaths" as the *Death Rate per 100,000 in the past 7 days* (ending Oct. 30th). Total deaths is inappropriate, as the population per state varies widely, skewing the data in largely populated states. The total death rate per 100,000 is also inappropriate, as many states had large first waves in the beginning of the year which will skew the current state of the pandemic. Using the current data (the past 7 days) will give a good analysis if the mandates are working *currently*.

We will operationalize "mandates" in two ways, as TRUE/FALSE indicators. One will be if the state mandates for their citizens to wear masks in public the other is if the states have issued mandates for employees who interact with the public to wear masks.

As we move through the model, we will do some descriptive analysis on other variables that may interact with the main question. These variables include: population density, racial diversity, political leanings, mobility, etc. We will analyze each of these variables to come up with a final model to determine if and how much face masks are currently reducing death rates of citizens due to COVID-19.

Import the data

```
library(magrittr)
library(tidyverse)
library(ggplot2)
library(readxl)
library(openxlsx)
library(stargazer)
library(lmtest)
library(sandwich)
```

```
library(patchwork)
covid_wb <- createWorkbook()
addWorksheet(covid_wb, "Covid-19")
raw_covid_data = read_excel("covid-19.xlsx", sheet = "Covid-19")
writeData(covid_wb, sheet = 1, x = raw_covid_data, startCol = 1, startRow = 0, colNames = TRUE)
saveWorkbook(covid_wb, "COVID_CLEAN.xlsx", overwrite = TRUE)
covid_data = read_excel("COVID_CLEAN.xlsx", sheet = 1)
file.remove("COVID_CLEAN.xlsx")
```

```
## [1] TRUE
```

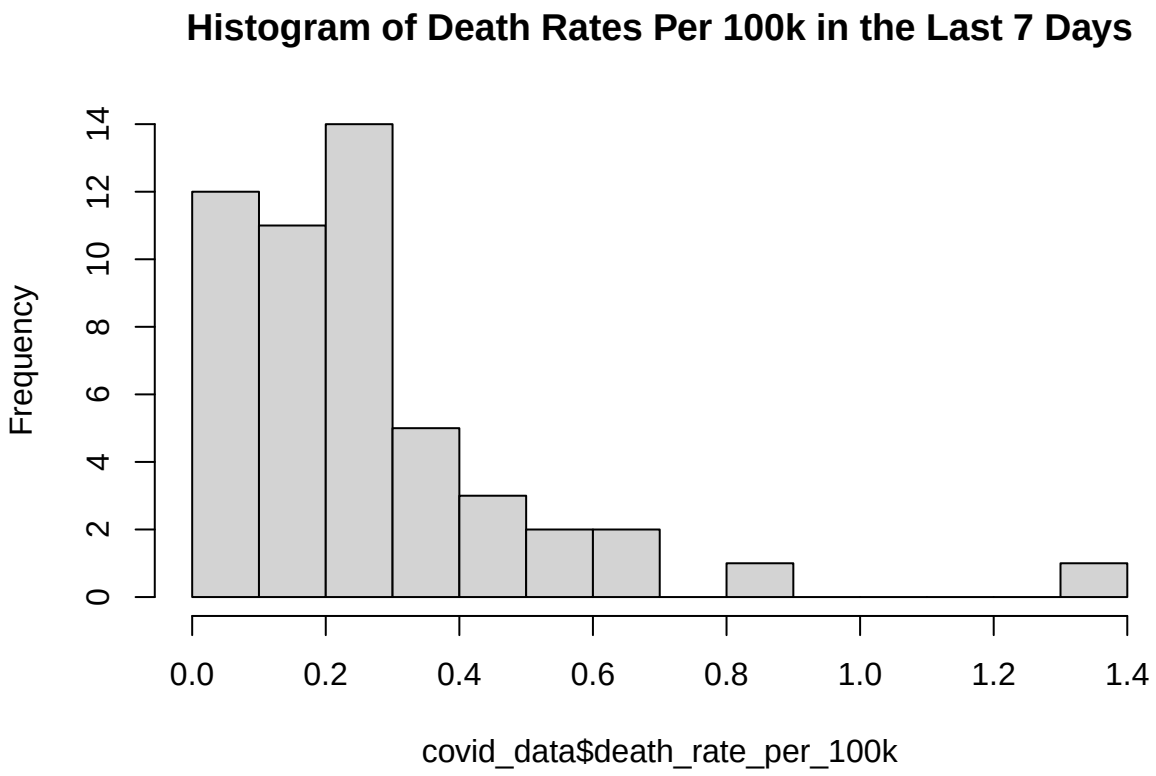
```
raw_covid_policy_data = read_excel("COVID-19 US state policy database (CUSP).xlsx",
                                   sheet = "Face Masks", range = "A1:J52")
covid_data <- left_join ( covid_data, raw_covid_policy_data, by=c("State"))
```

Initial Exploratory Data Analysis (EDA)

First, we want to take a look at the first variable we want to describe *Death Rate per 100K in the Last 7 Days*

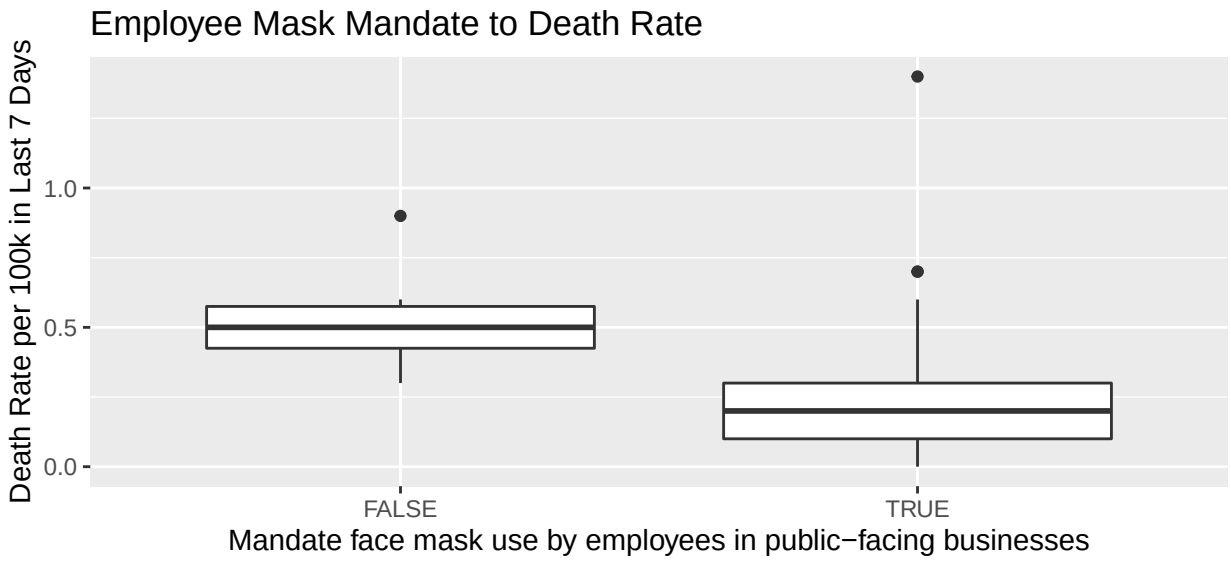
```
#make numeric
covid_data$death_rate_per_100k <- as.numeric(covid_data$`Death Rate per 100K in Last 7 Days`)

hist(covid_data$death_rate_per_100k, breaks = 15,
     main = "Histogram of Death Rates Per 100k in the Last 7 Days")
```



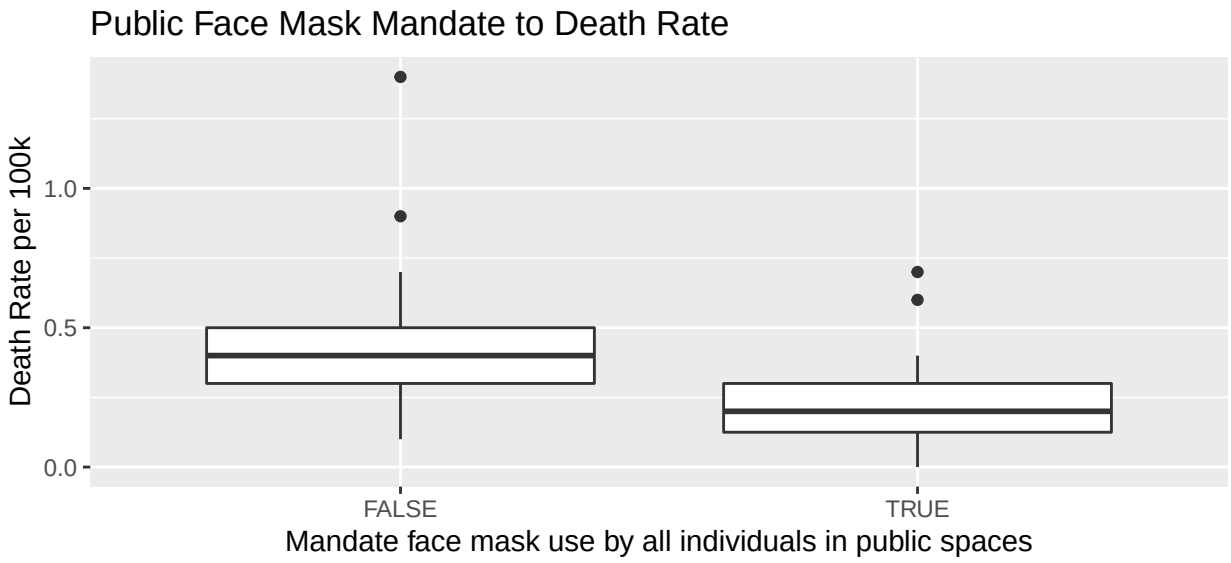
The distribution for "death rate" is skewed to the left. – *Need more discussion here*

```
covid_data$mask_mandate <-
  ifelse(covid_data$`Mandate face mask use by employees in public-facing businesses.y`
         == 0, FALSE, TRUE)
covid_data %>%
  ggplot(aes(x = mask_mandate, y = death_rate_per_100k)) +
  geom_boxplot() +
  labs(
    title = 'Employee Mask Mandate to Death Rate',
    x = 'Mandate face mask use by employees in public-facing businesses',
    y = 'Death Rate per 100k in Last 7 Days'
  )
```



Here there seems to be a difference between states that do have a face mask mandate for employees and those who don't. This seems significant enough to try in our initial model.

```
covid_data$mask_mandate_public <-
  ifelse(covid_data$`Mandate face mask use by all individuals in public spaces.y` == 0,
         FALSE, TRUE)
covid_data %>%
  ggplot(aes(x = mask_mandate_public, y = death_rate_per_100k)) +
  geom_boxplot() +
  labs(
    title = 'Public Face Mask Mandate to Death Rate',
    x = 'Mandate face mask use by all individuals in public spaces',
    y = 'Death Rate per 100k'
  )
```



The same can be said for the Public Face Mask Mandate. Let's see if we operationalize these together:

```
covid_data$mask_mandates <-
  ifelse(covid_data$`Mandate face mask use by all individuals in public spaces.y` == 0,
  ifelse(covid_data$`Mandate face mask use by employees in public-facing businesses.y`
         == 0, FALSE, TRUE), TRUE)

covid_data %>%
  ggplot(aes(x = mask_mandates, y = death_rate_per_100k)) +
  geom_boxplot() +
  labs(
    title = 'Any Face Mask Mandate to Death Rate',
    x = 'State either has a Public Mandate or Employee Mandate',
    y = 'Death Rate per 100k'
  )
```

We will use these two variables for our first model.

Model 1

Our first model will use the variables we looked at in the initial EDA.

$$death\_rate = \beta_0 + \beta_1 mandate + w \tag{1}$$

```
model1 <- lm(death_rate_per_100k ~ mask_mandates, data = covid_data)
coeftest(model1, vcov = vcovHC)
```

```
##
## t test of coefficients:
##
##                     Estimate Std. Error t value  Pr(>|t|)
## (Intercept)         0.533333   0.092376  5.7735 5.217e-07 ***
## mask_mandatesTRUE -0.253333   0.098785 -2.5645   0.01345 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Here we can see the coefficient of the mask_mandates is significant, but not highly, with the p-value at 0.01345. Practically speaking, it says that having either type of a face mask mandate (public or employee) is associated with a reduction of 0.25 deaths per 100k people, or 2.5 deaths per 1 million people.