Final Report

Claire Carlson, Sydney Scalzo, Ary Baal, Manaswi Kolani

Introduction

The COVID-19 pandemic has had a profound impact on many communities across the U.S. where some were disproportionately affected more than others. In this paper, we look into the impacts that COVID-19 had on the U.S. inmate populations and focus on the proportion of COVID-19 inmate deaths. More specifically, we focus on the proportion of COVID-19 inmate deaths in the state of Wisconsin and compare it to that of the average proportion COVID-19 inmate deaths across the U.S.

The goal of this report is to determine whether the state of Wisconsin's inmate population was disproportionately affected by COVID-19 relative to the U.S. inmate populations as a whole. By running a formal analysis of the difference in proportions between the proportion of Wisconsin's COVID-19 inmate deaths and the average proportion of COVID-19 inmate deaths across the U.S., this report attempts to explore the possibility of COVID-19 having a disproportionate impact on Wisconsin inmate population.

Through this analysis, we attempt to draw a conclusion about the impact that COVID-19 might have had on prisoners in Wisconsin in relation to the U.S. inmate population as a whole. Is the proportion of inmate deaths due COVID-19 in the state of Wisconsin the same as the average across the U.S.? The answer is no, as shown by the difference in proportions hypothesis test, there is a statistically significant difference in the proportion of inmates from Wisconsin that died from COVID compared to the proportion of inmates in the US who died from COVID.

Background

The data were collected by the New York Times in "facilities.csv" from many places, including websites overseen by state and federal prison systems and Immigration and Customs Enforcement, direct inquiries, public records requests, news conferences, meetings of county or state officials, coroner's reports, medical records, and reports from investigative agencies.[1]

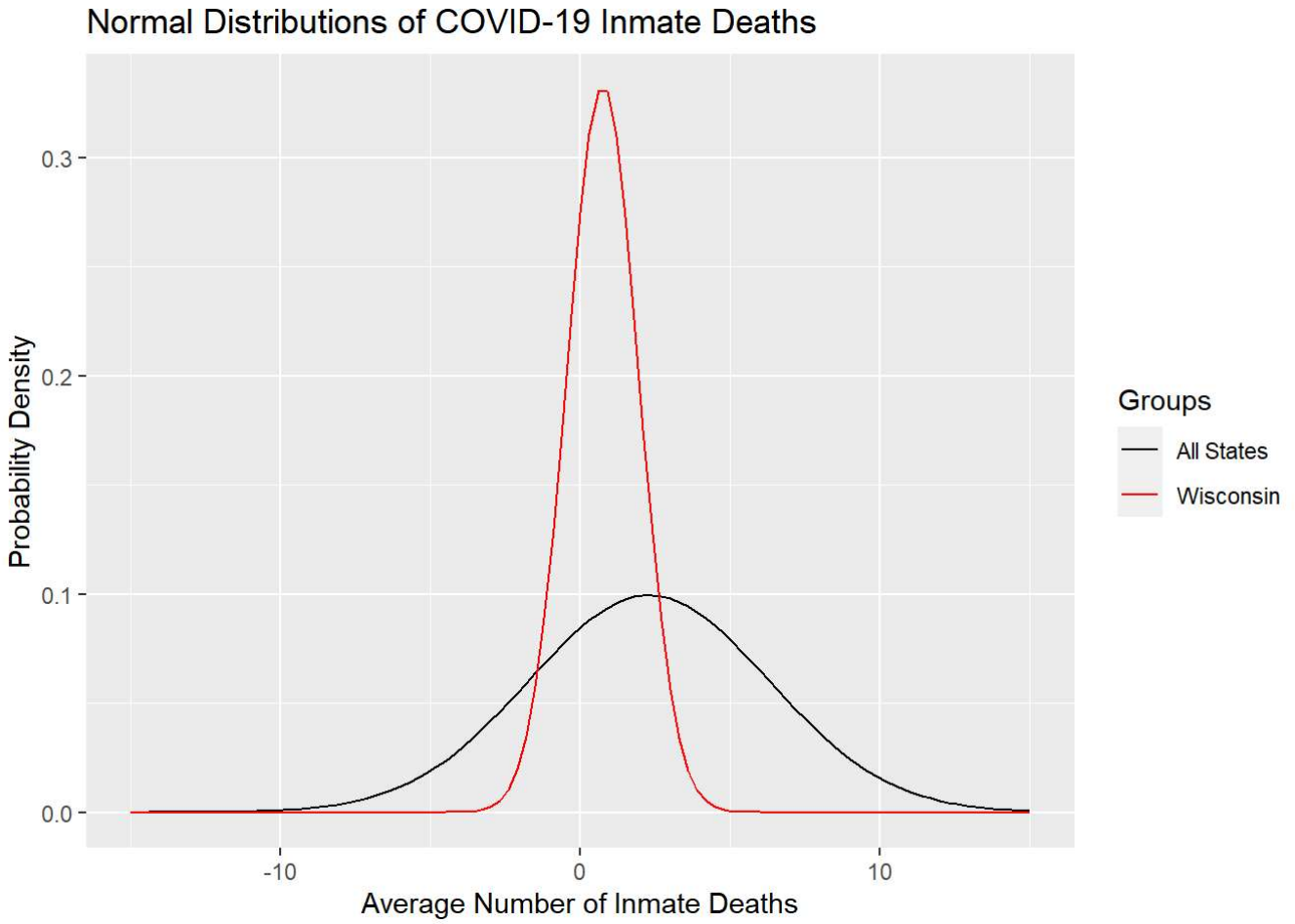
The key variables of the dataset are facility_state, which records the state each facility is located in, max_inmate_population_2020, which is the maximum facility population recorded between March 2020 and March 2021, and total_inmate_deaths, which is the number of inmates reported to have died of COVID-19 from the beginning of the pandemic through March 2021.

The background for this question is that in March 2020, the world was first put in lockdown for the novel Coronavirus-19 pandemic.[2] People were expected to stay six feet away from each other during this time to reduce the spread of the virus. Prisons were impacted by COVID negatively because of the close quarters which most inmates live in, and the overcrowding of some facilities. The data will help us examine how the pandemic affected each state in terms of how many inmates died from the coronavirus.

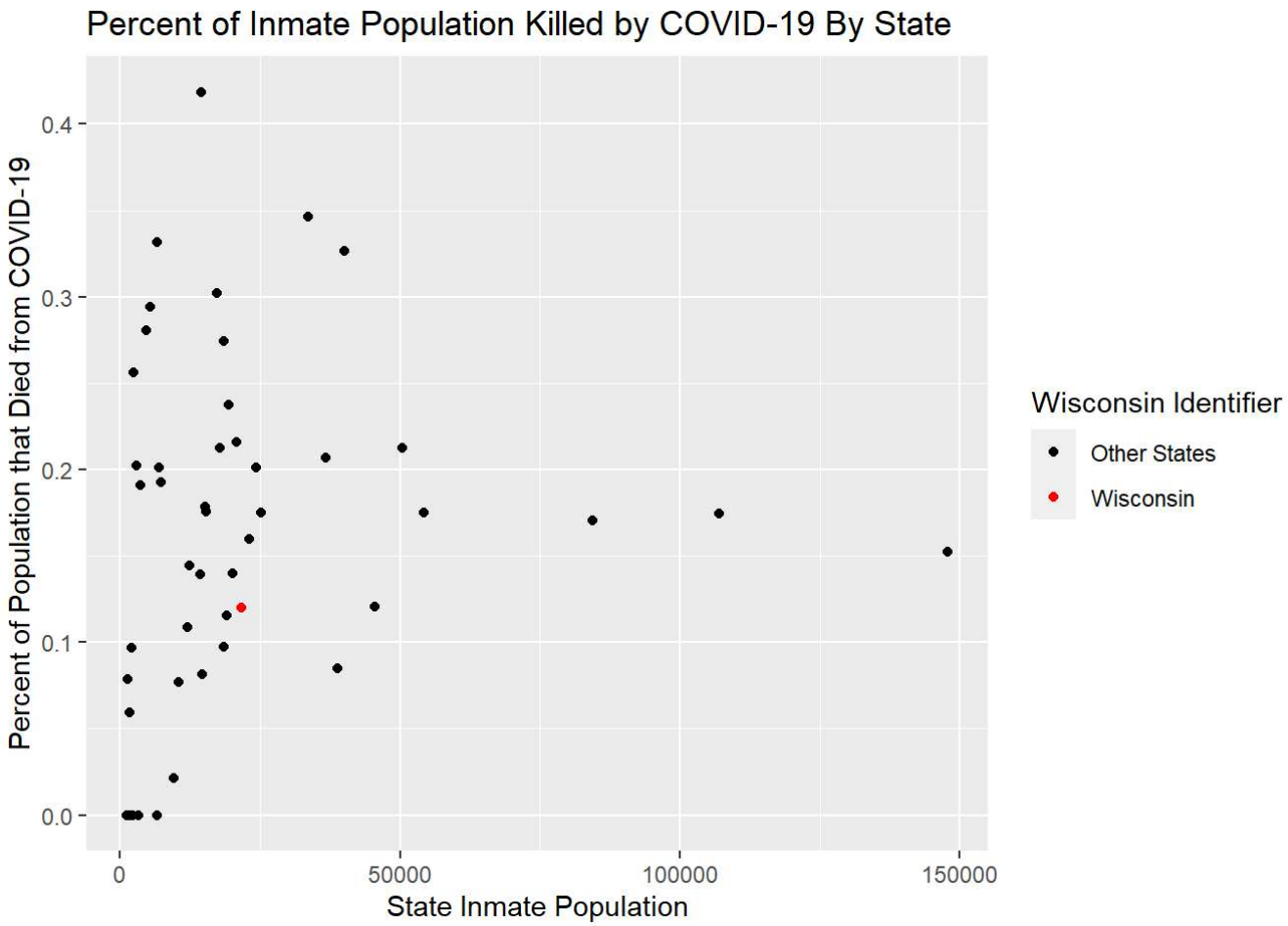
Unusual factors which may affect the interpretations of the results are different states releasing certain populations of prisoners in order to reduce their inmate population. Once the country knew that this virus was spreading through the air, certain states were pushing and trying to get as many people out of prison as possible, who were not a threat to society, for the safety and health of the prison population. Another unusal factor that may affect our interpretation is the fact that some of the data seems to be low/missing. West Virginia, Alaska, Idaho, Maine, Nebraska, Puerto Rico, and Vermont have 0 reported deaths from their prison population, which could be due to low/no testing going on, or refusal to share COVID data with reporters. Whatever the case, we will keep these factors in mind when interpreting our results.

For the rest of this report, we will analyze whether Wisconsin, our state of interest, has a different proportion of inmates that died from COVID-19 than the rest of the US inmate population proportion of COVID deaths. We will do this by performing a difference in proportions hypothesis test, and approximating with a normal distribution. We are testing the hypothesis that the proportion of deaths from COVID in prisons between Wisconsin and the US are not different. Based on the significance value, we will draw conclusions about what could lead to this outcome in our discussion.

Graphs



The graph above depicts the normal distribution of the average number of COVID-19 deaths of inmates in just Wisconsin as well as in all states across the US. The peak of each normal distribution curve corresponds to the average number of inmate deaths for that distribution. The graph shows the peak for the Wisconsin distribution (in red) is shifted left relative to the peak of the distribution for all states (in black), suggesting the average number of inmate deaths in Wisconsin is different from the average inmate deaths for the whole country.



The above scatter plot shows the relationship between the total inmate population of a state and the percentage of that population that died from COVID-19. Each point represents a unique state, with Wisconsin shown in a different color. It is important to keep in mind the varying sizes of populations in different states when comparing the rate of inmate COVID-19 deaths. For instance, a state with an extremely high population and a low death rate should be distinguished from another state with the same low death rate but a much smaller population.

Analysis

For our analysis we will use a difference in proportions hypothesis test to prove our thesis.

Statistical Model

The statistical model is:

- X_1 is our testing sample of inmates that will die from COVID in the US
- X_2 is our testing sample of inmates that will die from COVID in the state of Wisconsin
- p_1 is the probability that an inmate dies from COVID in the US
- p_2 is the probability that an inmate in WI dies from COVID

$$X_1 \mid p_1 \sim \text{Binomial}(1063866, p_1)$$
$$X_2 \mid p_2 \sim \text{Binomial}(21645, p_2)$$

Hypothesis

- Our hypothesis is that the probability of an inmate dying from COVID in Wisconsin's inmate population is no different than the probability of an inmate dying from COVID in the whole US inmate population.

$$H_0 : p1 - p2 = 0$$
$$H_a : p1 - p2 \neq 0$$

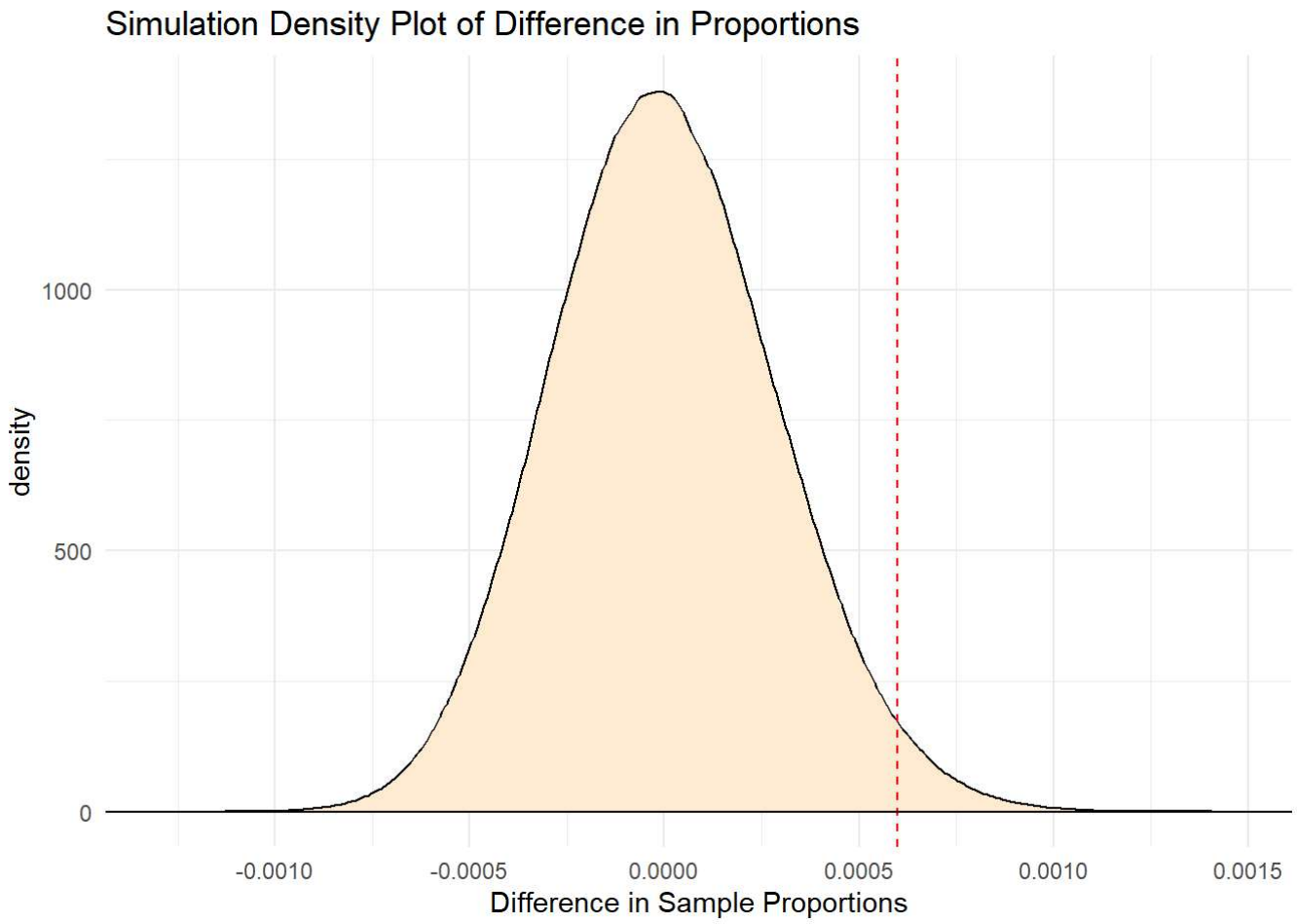
- If the null hypothesis is true, then p1=p2, the distribution of the test statistic is whatever it is when X1 and X2 are drawn from binomial distributions with the same success probability p

Test Statistic

- For this test statistic, we are calculating a p-hat value using our proportion estimates (p1 and p2) from the data in a simulation.
- We are simulating many independent samples of the same sample size, then calculating the difference in these sample proportions.
- By simulating these proportions we can use this as the long term average estimate and draw conclusions about our hypothesis
- Our test statistic is

```
## [1] 0.0005988375
```

Sampling Distribution



Outcomes

- We see that the simulation sampling distribution is well-approximated by a normal curve
- the test statistic is far in the tail, so we expect a small p-value

P-Value

- Our p-value is

```
## [1] 0.038498
```

- We are testing the null hypothesis against a significance values of 0.05.
- The p value is 0.03 which indicates strong evidence against the null hypothesis, and we can reject that there are equal proportions of death from COVID among Wisconsin inmates compared to all US inmates.

Interpretations

With respect to our question of interest, we can draw a conclusion based on the results of our analysis that the two proportions are likely not equal. This conclusion is also supported by a scatter plot of our data which shows that Wisconsin's proportion of COVID inmate deaths is lower than that of other states with similar populations. That said, we are aware of the limitations of our analysis. A two-sided test doesn't tell us much in regard to the direction of impact–we do know that the proportion of inmate deaths for the state of WI is statistically significant, but we don't know whether that is a desirable outcome. Did the state of Wisconsin do better than the rest of the U.S. in terms of lowering COVID inmate deaths…or did it fail to do that and instead had a higher COVID inmate death rate? Although we attempt to answer this question with graphical representation, we believe that conducting a one-sided test of significance in this case is more appropriate.

Also, due to the fact that data collection has stalled in many places, there could be significant differences in COVID deaths that would better represent the long term average in proportion of COVID inmate deaths that are not represented in our data.

Our main conclusion is that the state of Wisconsin does have a statistically significant different proportion of COVID inmate deaths, compared to the US inmate population as a whole. This conclusion is drawn based on our p-value that presents strong evidence that this is the case. By graphically exploring the data, it looks like Wisconsin's long run average proportion of COVID inmate deaths is lower, but we cannot make that conclusion definitively. We speculate that part of the reason why the state of Wisconsin may have had lower proportion of COVID inmate deaths was due to the fact that Wisconsin's DOC implemented rigorous COVID policies to fight the spread as soon as the pandemic started. [3]

---

1. https://github.com/nytimes/covid-19-data/tree/master/prisons (https://github.com/nytimes/covid-19-data/tree/master/prisons)↩
2. https://en.wikipedia.org/wiki/COVID-19_pandemic (https://en.wikipedia.org/wiki/COVID-19_pandemic)↩
3. https://doc.wi.gov/Pages/COVID19(Coronavirus)/ResponseEfforts.aspx (https://doc.wi.gov/Pages/COVID19(Coronavirus)/ResponseEfforts.aspx)↩