Research Proposal: How are state policies and population related to the state-wide spread of COVID-19?

Cody Burker, Emily Fernandes, Margo Suryanaga

Contents

1 Background

The COVID-19 pandemic is the first pandemic seen in a century, affecting each individual worldwide. While many countries issued lockdowns, travel quarantines, and other COVID-19 restrictions on a federal level, the US federal government placed the power with the governors to regulate the implementation of these restrictions on a state-by-state basis.

There has been increased skepticism on the true effectiveness of restrictions in slowing down the spread of COVID-19, especially due to the amount of economic impact resulting from these procedures. Evaluating how each state's population density, lockdown procedures, and facemask policies correlates to the spread of COVID-19 within that state helps citizens understand the importance of following these policies. Understanding the effectiveness of COVID-19 restrictions in slowing the spread of the disease also helps policy makers justify the implementation of these policies to their constituents. Additionally, this information might be of interest to public health officials who are looking for best practices to contain a future pandemic or infectious disease.

2 Data sources

1. COVID-19 US State Policy Database A database of state policy responses to the pandemic, compiled by researchers at the Boston University School of Public Health.
2. NY Times Covid-19 Data Repository A series of data files with cumulative counts of coronavirus cases in the United States, at the state and county level, over time.

From data source (1), we expect to get:
- Closure and reopenings: Start and end dates of closures of public spaces
- Stay at home: Start and end dates of stay at home policies
- Face masks: Start and end date of face mask mandates
From data source (2) we expect to get:
- Weekly COVID-19 case counts

3 Methodology and Expected outcome

Our team would like to study the initial spread of the Covid-19 virus, from the first case seen in the United States until May 31st. Our goal was to create a series of models to explain the relationship between the various state policies put in place to curtail the spread of the virus and the number of cases observed in each state. In the analysis we will use the following policies as input variables: *Business mask mandate (measured in days the policy was in place)* Stay at home/shelter in place order (measured in days the policy was in place) *Interstate travel quarantines (measured in days the policy was in place)* Restaurant/Bar/Gym Shutdowns (measured in days the policy was in place) *Shutdown declared in relationship to day* State of Emergency declared in relationship to 3-11-2020 (Date that the World Health Organization declared the Covid-19 Pandemic) relationship to 3-11-2020 (Date that the World Health Organization declared the Covid-19 Pandemic)

We understand that the spread of Covid-19 is also influenced by not just state policies, but also by the relative closeness of people within each state. In an effort to capture this, we will also include the population density as an input variable. With these inputs, we hope to explain which state policies were effective at curtailing the initial spread of the Covid-19 virus. For each model we will evaluate the sssumptions for a Classical Linera Model. These requirements are:

1. I.I.D.
2. Linear Conditional Expectation Exists
3. No Colinearity
4. No Homoskedastic Error
5. Normal Residuals

The data use for the following models

First, we began to explore the number of cases in each state per 100,000 residents. This allowed us to account for the number of cases while equalizing the predicted variable to account for the fact that New York or California will have a much larger number of cases than Alaska simple due to the fact these states have many more residents. After viewing the histogram of the number of cases per state per 100k residents, our team chose to handle the skew of the data by completeing a log transformation. The histogram of the log transformed variable can be seen below.

Predicted Variable Histograms
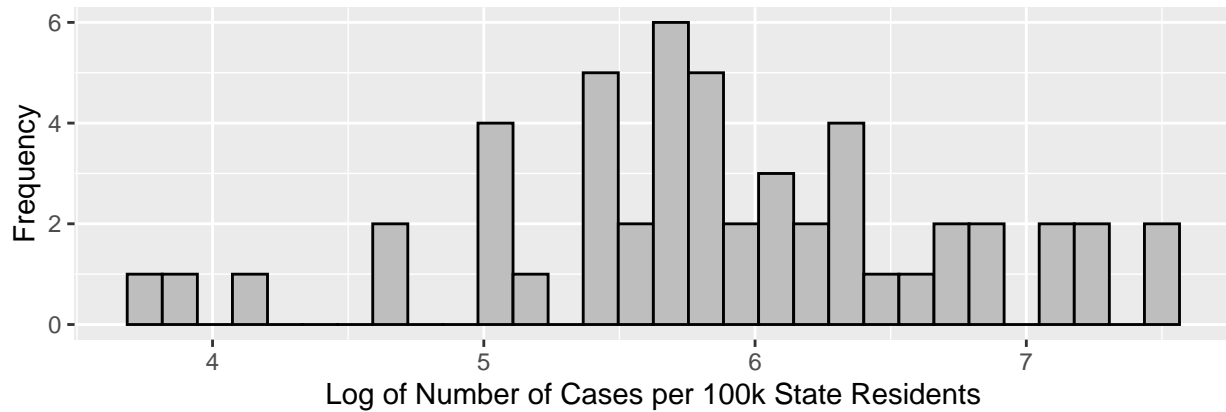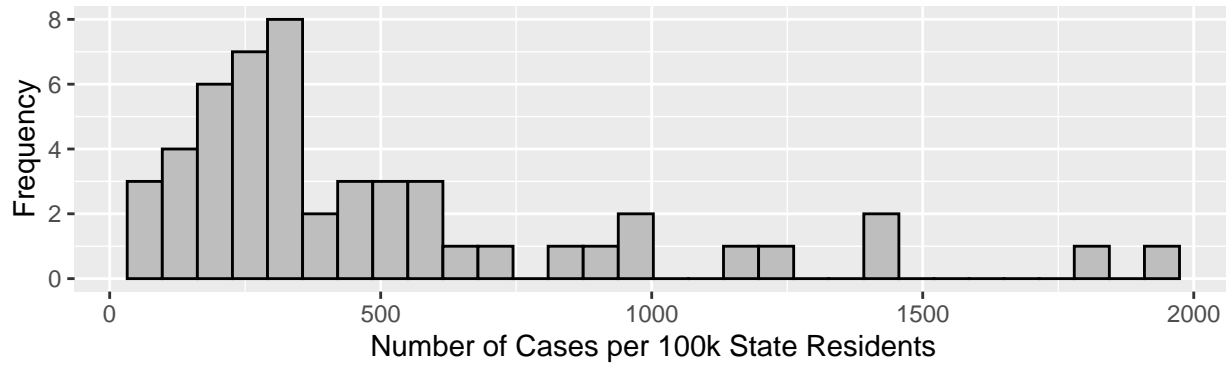
For our base model, we chose to select the one input variable that was a characteristic of the state and not a policy enacted by the state. We hypothesize that the state's population density is one of the first building blocks to describe the number of cases in that state on May 31st, 2020. Close human contact is how this virus spreads and population density captures the relative closeness of residents in a state. Again, after viewing the initial highly skewed histogram we chose to complete a log transformation of this variable. The transformed histogram can be seen below.

Input Variable Histograms
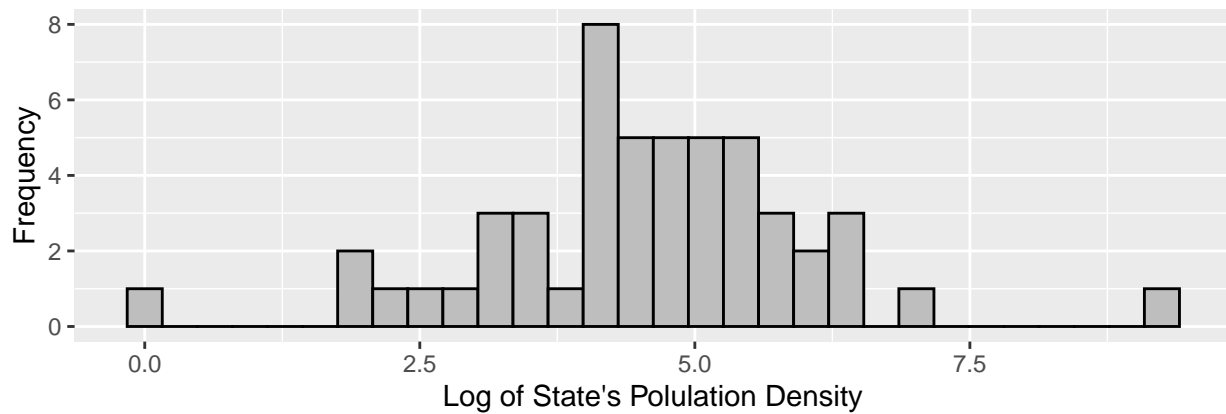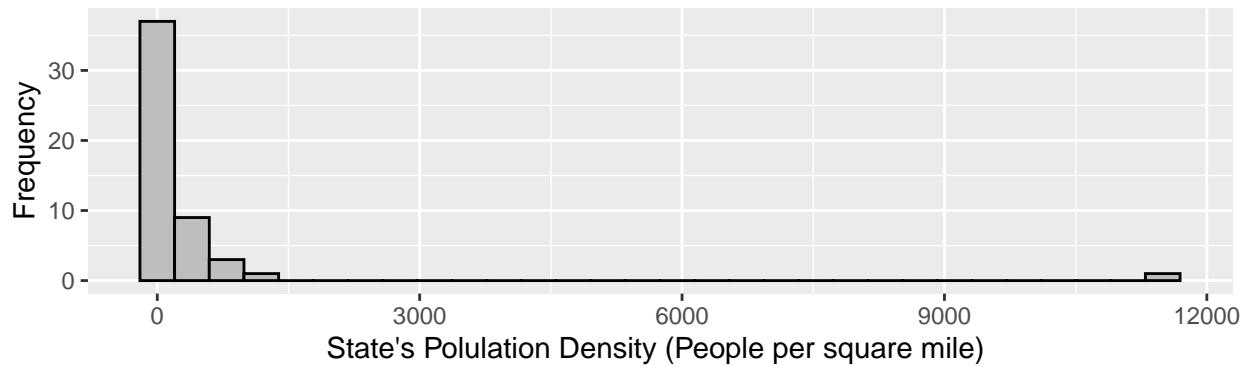

Our base model will look at the log relationship of each state's cases on May 31st, 2020 per 100k residents with

Relationship of Log(Covid−19 Cases per 100k Residents) and Log(Population Density)
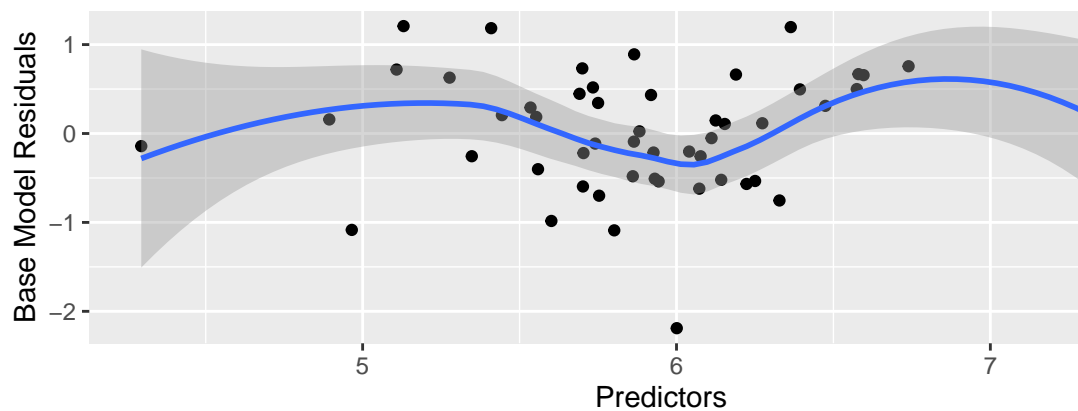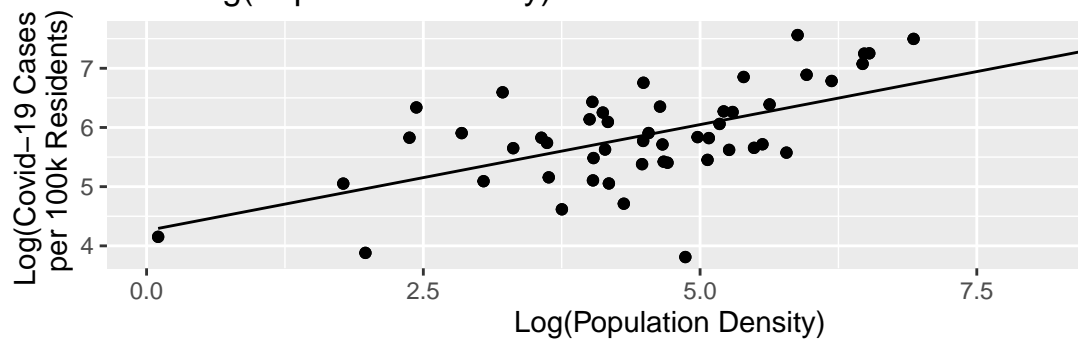
that state's population density.

From the Residual/Predictors plot shown above, one can see that this basic model's predictors has a fairly linear relationship with residuals implying that a Linear Conditional Expectation exists. Colinearlity cannot exist for this base model because we only have 1 input variable. The Residual/Predictors plot displays a flaring of the standard error at the extremes which implies the potential for Homoskedastic errors, therefore we will use robust standard error when reviewing the significance level of this model's coefficients. Lastly we have a plot of the basic model's residuals to ensure they are normal. As you can see from the histogram below, the residual's are in fact normally distributed.
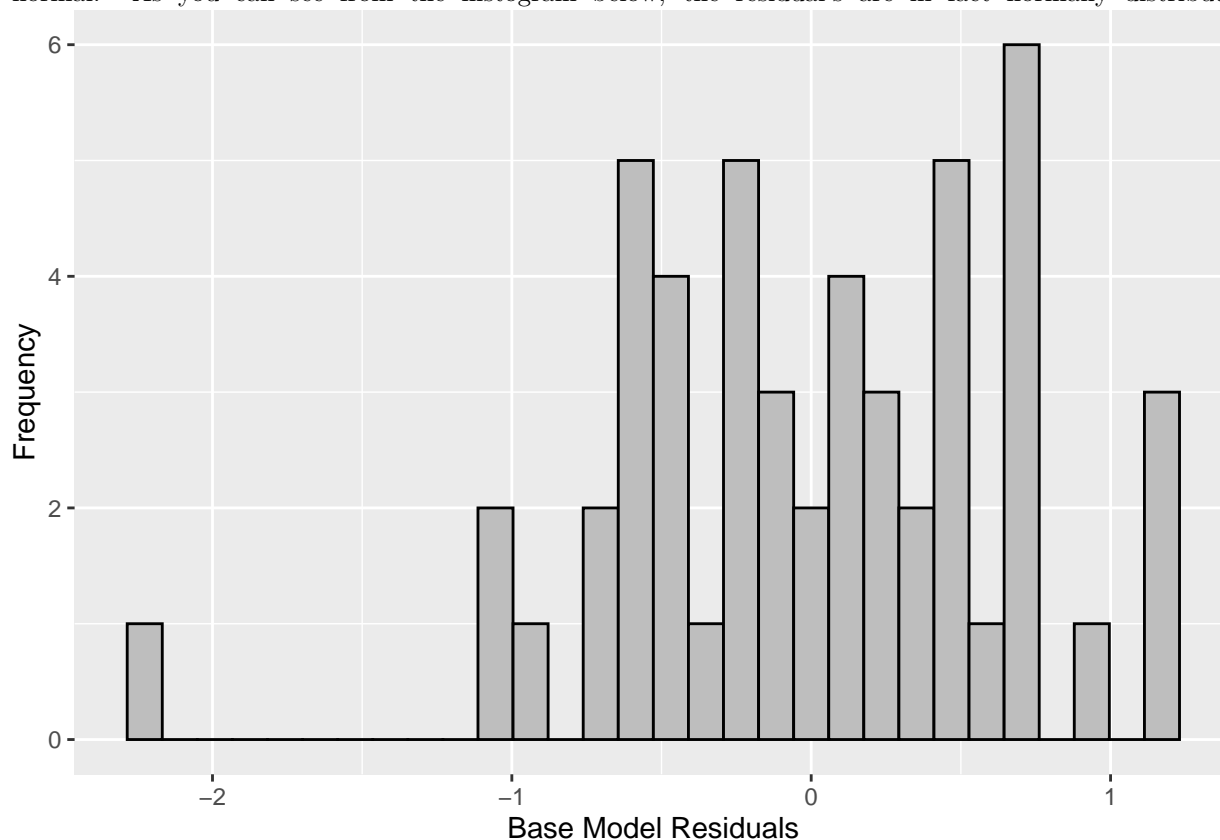
For our next model, we chose to build off the basic model while adding several inputs to represent key state policies. This next model will also account for the number of days each state had a: *Business mask mandate* Stay at home/shelter in place order *Interstate travel quarantine

We first

4 Regression Table

Table 1 displays the regression of all 3 models. We are use robust standard error.

4.1 Model 1

We will begin by inspecting the first Model in Table 1. Here we see a significant relationship between the Log (Population Density) and
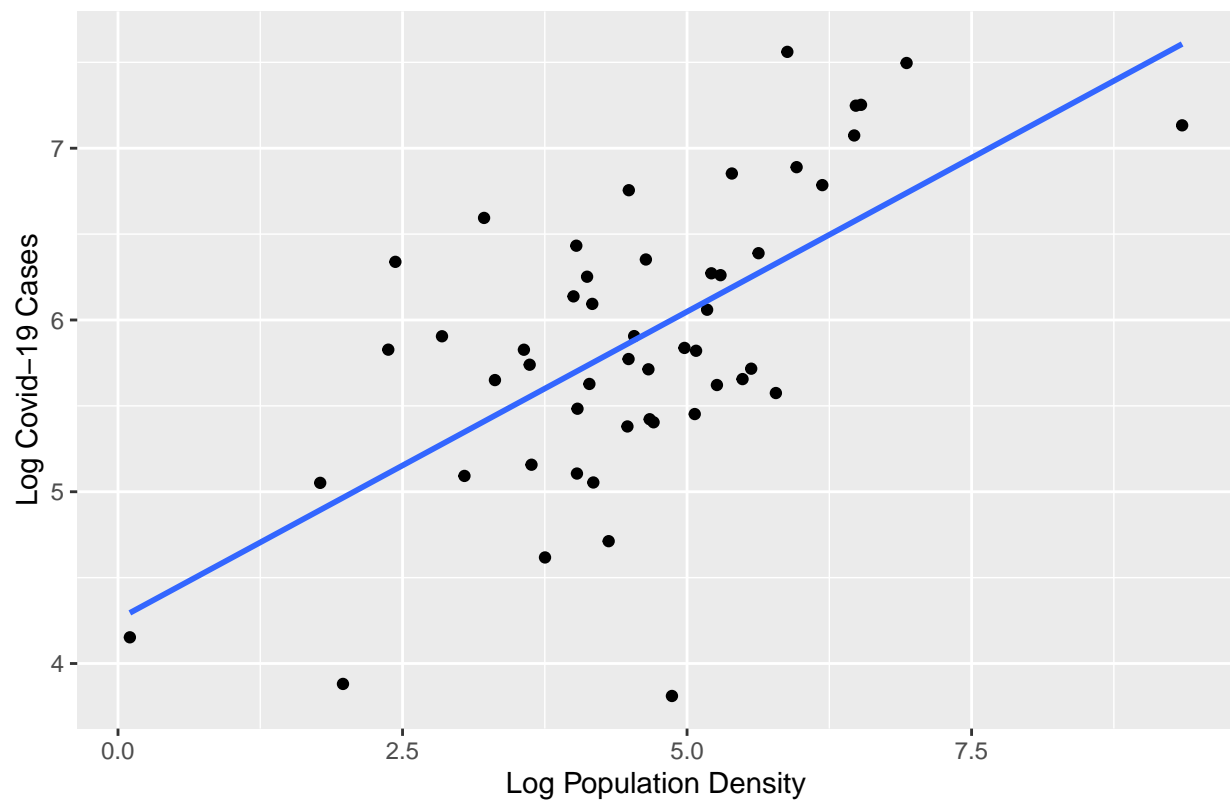
Table 1: Regression Table

	Dependent Variable: Log (Covid-19 Cases)		
	Simple Model	Moderate Model	Complex Model
	(1)	(2)	(3)
Log (Population Density)	0.358***	0.329***	0.350***
	(0.063)	(0.084)	(0.105)
Business mask mandate (days)		0.016***	0.016**
		(0.006)	(0.007)
Stay at home/shelter in place order (days)		−0.010**	−0.010
		(0.005)	(0.007)
Interstate Travel Quarantine In Place Length (days)		−0.009**	−0.009**
		(0.004)	(0.005)
Restaurant Shutdown Length (days)			−0.002
			(0.015)
Bar Shutdown Length (days)			−0.008
			(0.018)
Gym Shutdown Length (days)			0.011
			(0.016)
State of Emergency Speed (days)			−0.027
			(0.033)
Constant	4.257***	4.628***	4.587***
	(0.306)	(0.379)	(0.995)
F-Statistic	32.4*** (df = 1; 49)	32.4*** (df = 5; 45)	4.15*** (df = 9; 41)
Observations	51	51	51
R^2	0.391	0.586	0.601
Adjusted R^2	0.378	0.550	0.525
Residual Std. Error	0.675 (df = 49)	0.574 (df = 46)	0.590 (df = 42)

Note: $^*p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Model 1: Log Population Density vs Log Covid−19 Cases

4.2 Model 2

4.3 Model 3