

Data Analysis

Michael Lewis, Aarushi Sahejpal, Katherine Prince

2023-01-08

This document contains the code needed to conduct our statistical and descriptive analyses.

Data Import

```
kableExtra::kable(head(df))
```

```
## Warning in !is.null(rmarkdown::metadata$output) && rmarkdown::metadata$output
## %in% : 'length(x) = 2 > 1' in coercion to 'logical(1)'
```

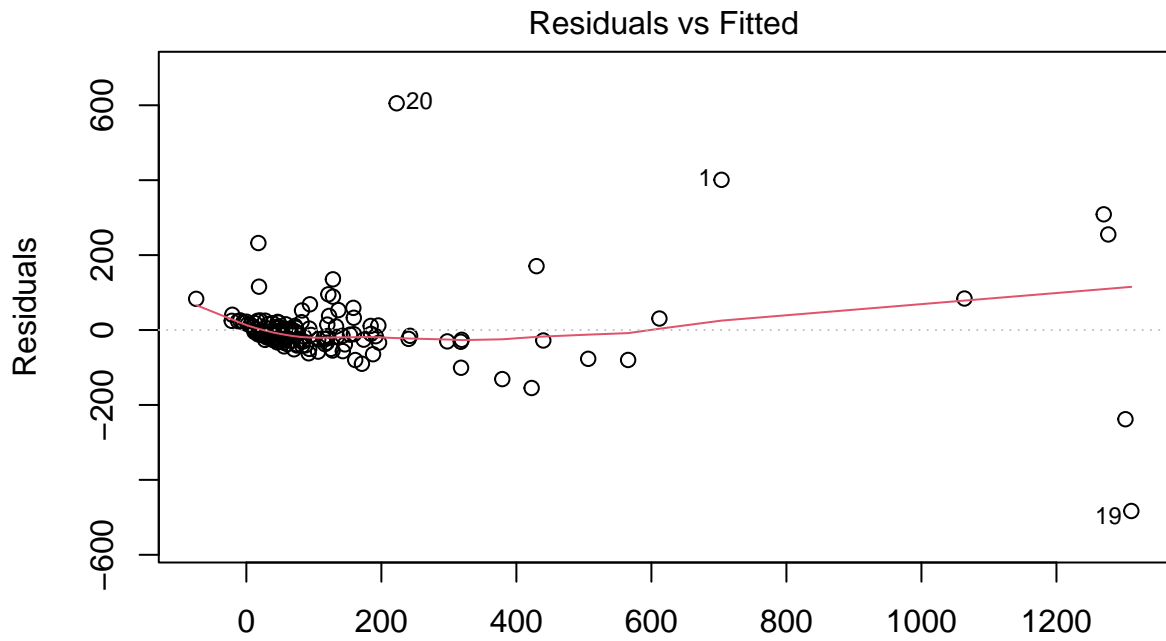
county	average_cases	average_deaths	geoid	state	estimate_number_commuting
District of Columbia	51847.198	1104.82967	11001	District of Columbia	
Allegany County	7694.165	217.65934	24001	Maryland	
Anne Arundel County	44252.050	642.59615	24003	Maryland	
Baltimore County	64934.606	1578.92857	24005	Maryland	
Calvert County	4488.251	82.92857	24009	Maryland	
Caroline County	2670.037	34.59341	24011	Maryland	

Linear Model

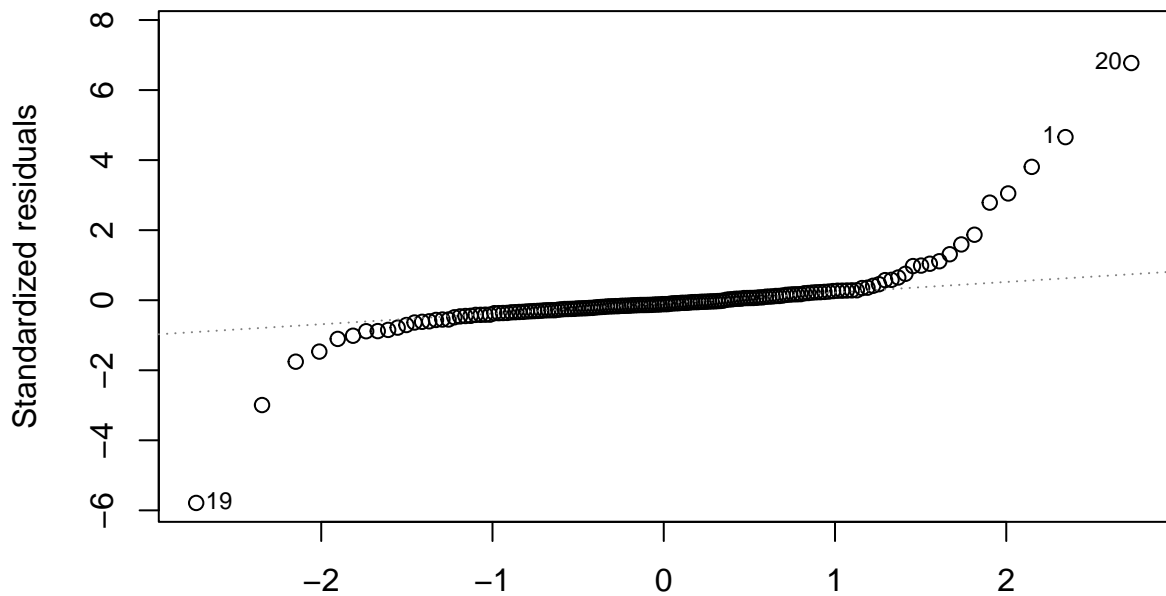
Table 1: Two Regression Models Predicting Variation in 2021 COVID Outcomes

	For all counties in DC, MD, VA	
	Average Cases	Average Deaths
Percent Essential Workers	105.619 (78.412)	1.862 (1.607)
Number of Individuals with Health Insurance	0.116*** (0.037)	0.005*** (0.001)
Number of Individuals Commuting to Work	-0.081 (0.071)	-0.008*** (0.001)
Mean Earnings (USD)	-0.028* (0.016)	-0.001** (0.0003)
Intercept	674.878 (2,535.645)	27.942 (51.976)
Corrected AIC	3113	1884.6
R ²	0.898	0.869
F Statistic (df = 4; 153)	337.657***	252.807***

Note: *p<0.1; **p<0.05; ***p<0.01



Fitted values
 $\text{lm}(\text{average_deaths} \sim \text{estimate_percent_essential_workers} + \text{estimate_health_in} \dots)$
 Normal Q-Q



Theoretical Quantiles
 $\text{lm}(\text{average_deaths} \sim \text{estimate_percent_essential_workers} + \text{estimate_health_in} \dots)$

