Linear Model 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

```
library(tidyverse)
library(stargazer)
library(AICcmodavg)
read_csv("../data_and_cleaning/covid_econ_data.csv") -> df
```

Linear Model

Overview

We begin by estimating an Ordinary Least Squares (OLS) model that predicts average cases and deaths (in 2021) for all counties in DC, Maryland, and Virginia. In our model, we considered: the percent of workers that could be considered essential, the number of individuals that have health insurance, the number of workers that commute to work, and the mean individual earnings. Since state and local vaccination efforts were underway for the duration of 2021, we also controlled for average vaccinations administered. The ultimate goal of our analysis was to ascertain which economic indicators were significant in predicting COVID outcomes.

Models

Overall, our models explain about 80-90% of variation in both average cases and deaths for all counties in DC, MD, VA in 2021. The following paragraphs will dive deeper into each model and elucidate the significance of our results.

Average Cases

Our baseline model explained about 89% of the variance in average cases, however, only three of the five predictors were statistically significant. To account for this, we estimated another linear model with just the statistically significant predictors (shown in Table 2 and below).

$$\widehat{Average\ Cases} = 164.7(Percent\ Essential\ Workers) + .039(Number\ of\ Individuals\ with\ Health\ Insurance)$$
$$+ .035(Average\ Vaccinations\ Administered) - 2691.9$$

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	128.566** (62.578)	2.880** (1.272)
Number of Individuals with Health Insurance	0.103** (0.042)	0.003*** (0.001)
Number of Individuals Commuting to Work	−0.121 (0.081)	−0.004** (0.002)
Mean Earnings (USD)	−0.015 (0.014)	−0.001* (0.0003)
Average Vaccinations Administered	0.037*** (0.004)	0.0003*** (0.0001)
Intercept	−887.164 (2,021.619)	−4.595 (41.079)
Corrected AIC	2611	1551.2
R^2	0.895	0.805
F Statistic (df = 5; 130)	221.441***	107.624***
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 2: Only Significant Variables for Predicting Variation in Average Cases

	For all counties in DC, MD, VA
	Average Cases
Percent Essential Workers	164.734*** (60.682)
Number of Individuals with Health Insurance	0.039*** (0.003)
Average Vaccinations Administered	0.035*** (0.004)
Intercept	−2,691.905* (1,509.005)
Corrected AIC	2611.1
R^2	0.891
F Statistic	361.109*** (df = 3; 132)
Note:	*p<0.1; **p<0.05; ***p<0.01

Notably, the percent of essential workers in a given county had by far the largest effect on average cases. Indeed, for every percent of workers in a county that could be classified as essential average cases will increase by about 164. The estimates for number of individuals with health insurance and average vaccinations administered were both much lower at .039 and .035, respectively. Despite being statistically significant, the estimates for these variables are so small that they are practically insignificant in the context of our analysis.

Average Deaths

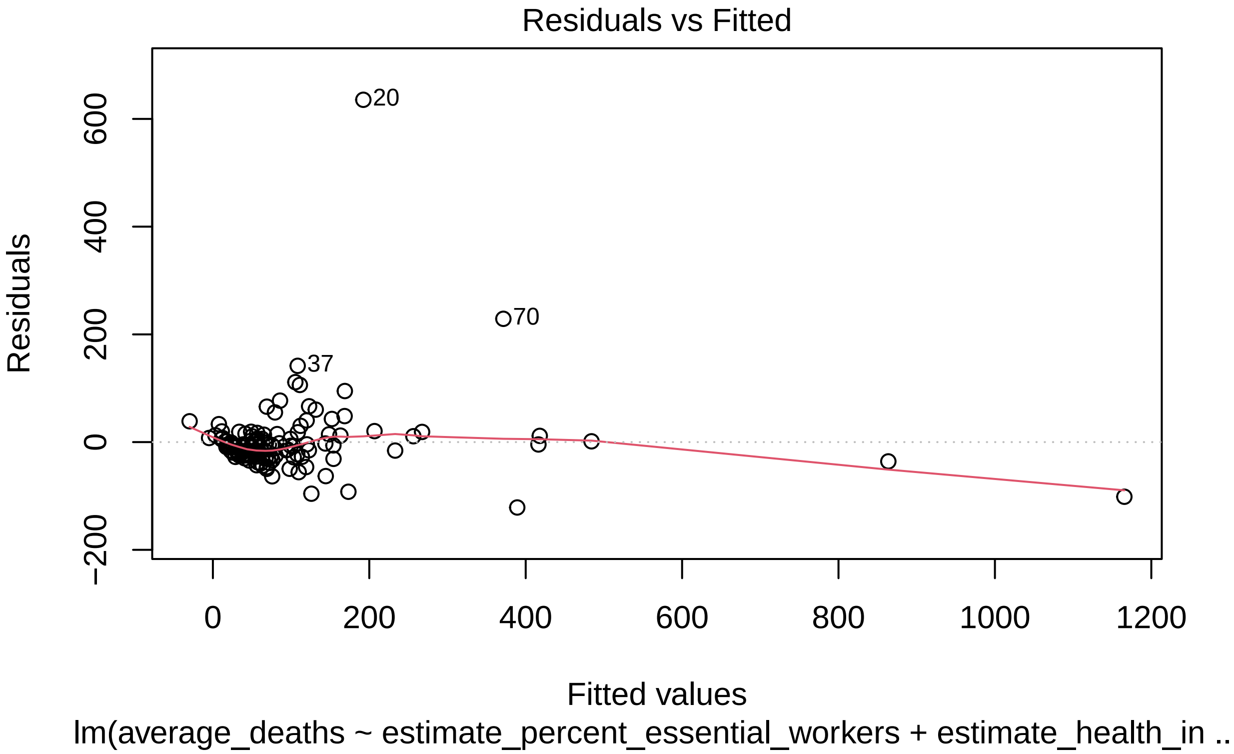
Our model to predict average deaths preformed better, with a comparable R-sq at .8 and lower AIC. Every predictor was statistically significant in explaining variation in average deaths. Notably, for every additional percent in essential workers we estimate average deaths to increase by about 2.8, while holding other variables constant. We also find that as mean earnings increases, we can expect average deaths to decrease. Specifically, an increase of $10,000 in mean earnings (for a given county) is predicted to result in a decrease of ten for mean deaths, while holding other predictors constant. While we cannot say this relationship is causal – it does highlight the importance of wealth and resources in impacting public health outcomes. Indeed, counties with higher earnings likely have better hospitals and public services, factors which one can reasonably assume have a positive impact on public health outcomes.

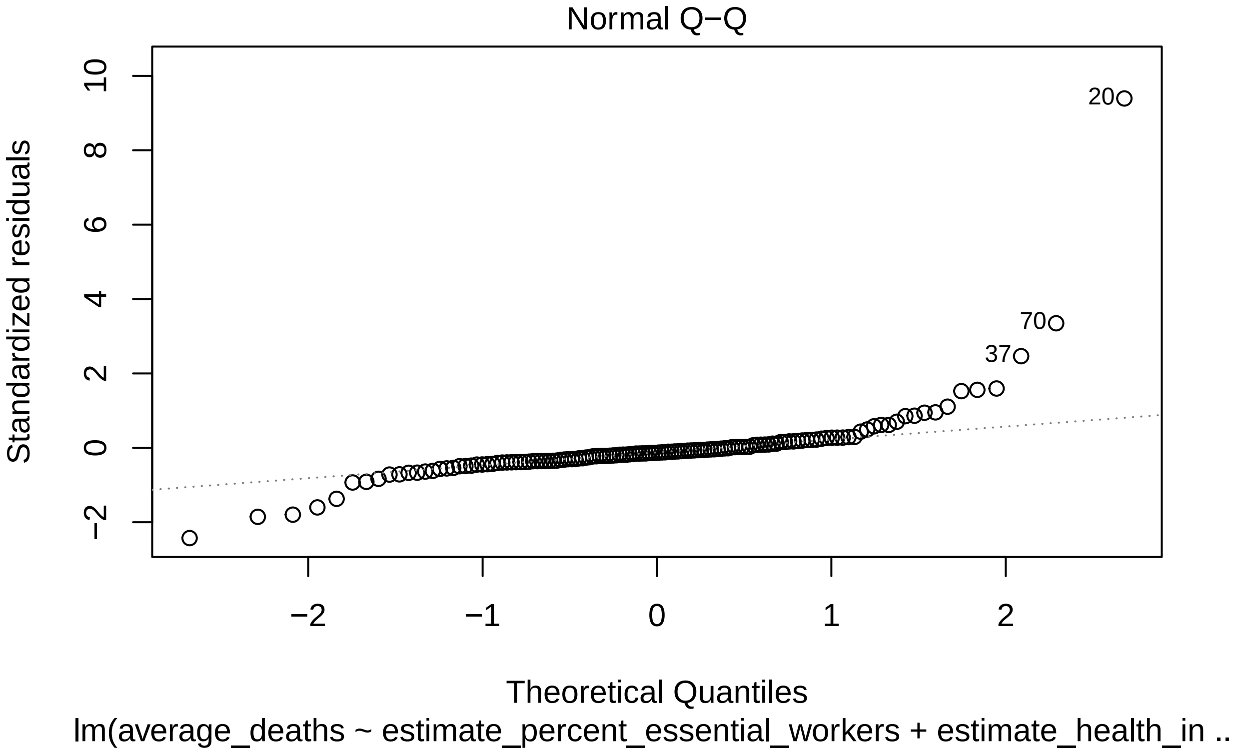
In our check for multicollinearity we found that the number of individuals with health insurance and number of individuals commuting to work are highly correlated as indicated by the VIF and Pearson's R-sq. Thus, we do not place much emphasis on the coefficients for these predictors. Indeed, while multicollinearity does make it difficult to disaggregate the effect of one predictor from another, it does not diminish the predictive power of a model.

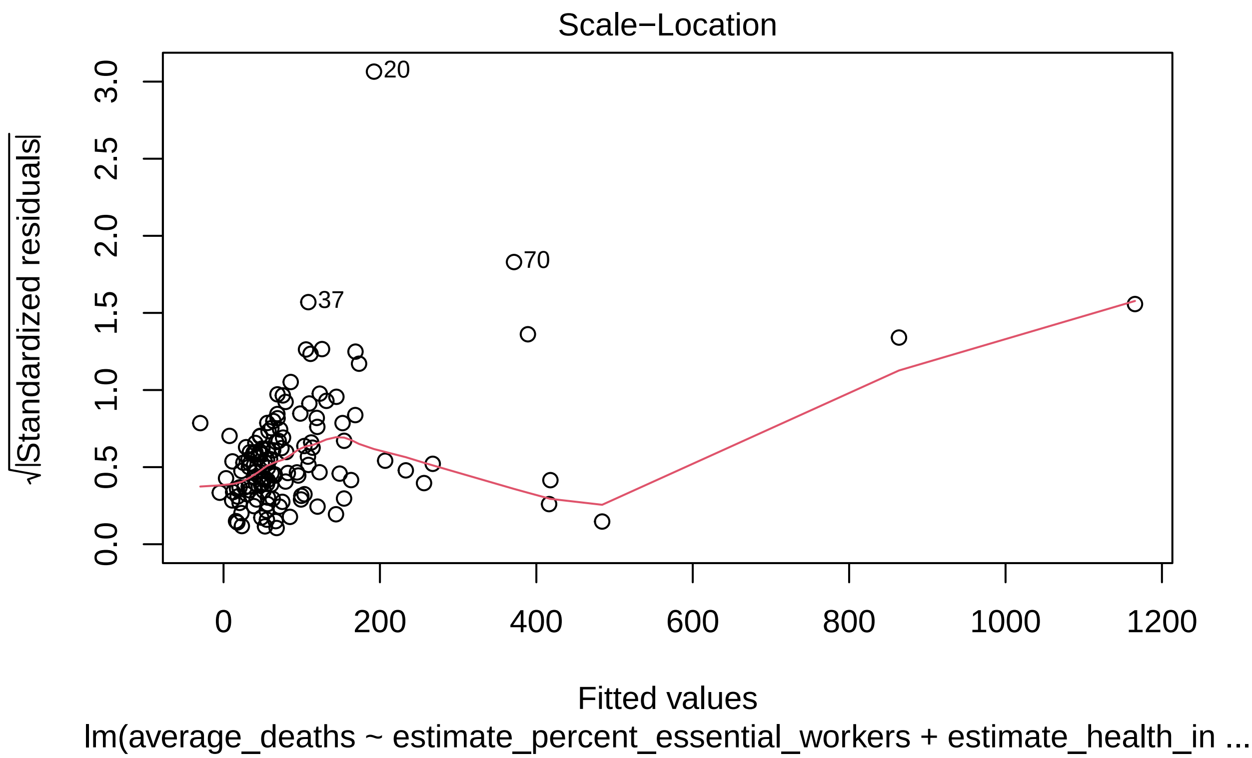
Finally, there is a positive relationship between average vaccinations administered and average deaths. As was the case before, this is probably due to endogeneity concerns. Vaccinations were increasing for much of 2021, however, at the same time COVID was still wreaking havoc on communities – and unfortunately resulting in lives being lost. This temporal aspect aside, many states/communities boosted or even prioritized vaccination efforts in areas that were particularly hard hit – another possible factor in explaining this result.

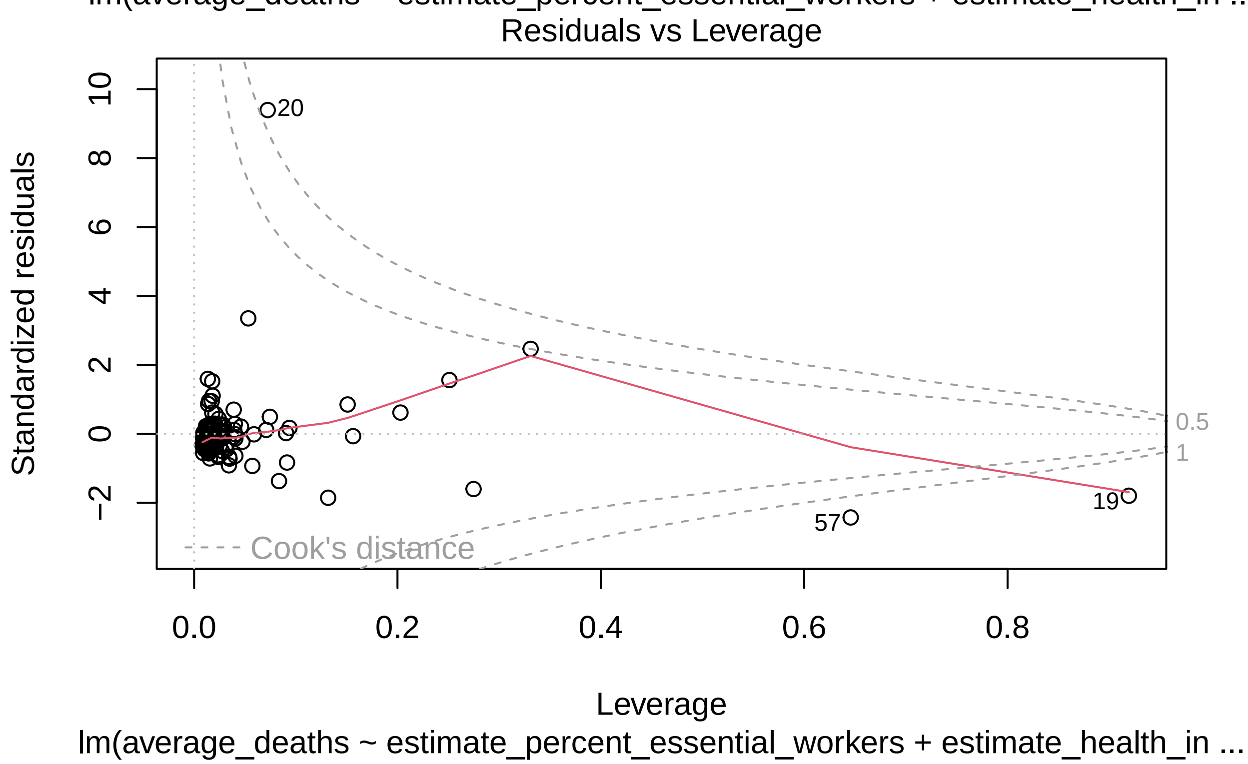
$$\begin{aligned}\widehat{Average\ Deaths} = 2.8(Percent\ Essential\ Workers) + .003(Number\ of\ Individuals\ with\ Health\ Insurance)\\ -.004(Number\ of\ Individuals\ Commuting\ to\ Work) - .001(MeanEarnings)\\ +.0003(Average\ Vaccinations\ Administered) - 4.59\end{aligned}$$

```
plot(average_death.full.lm)
```

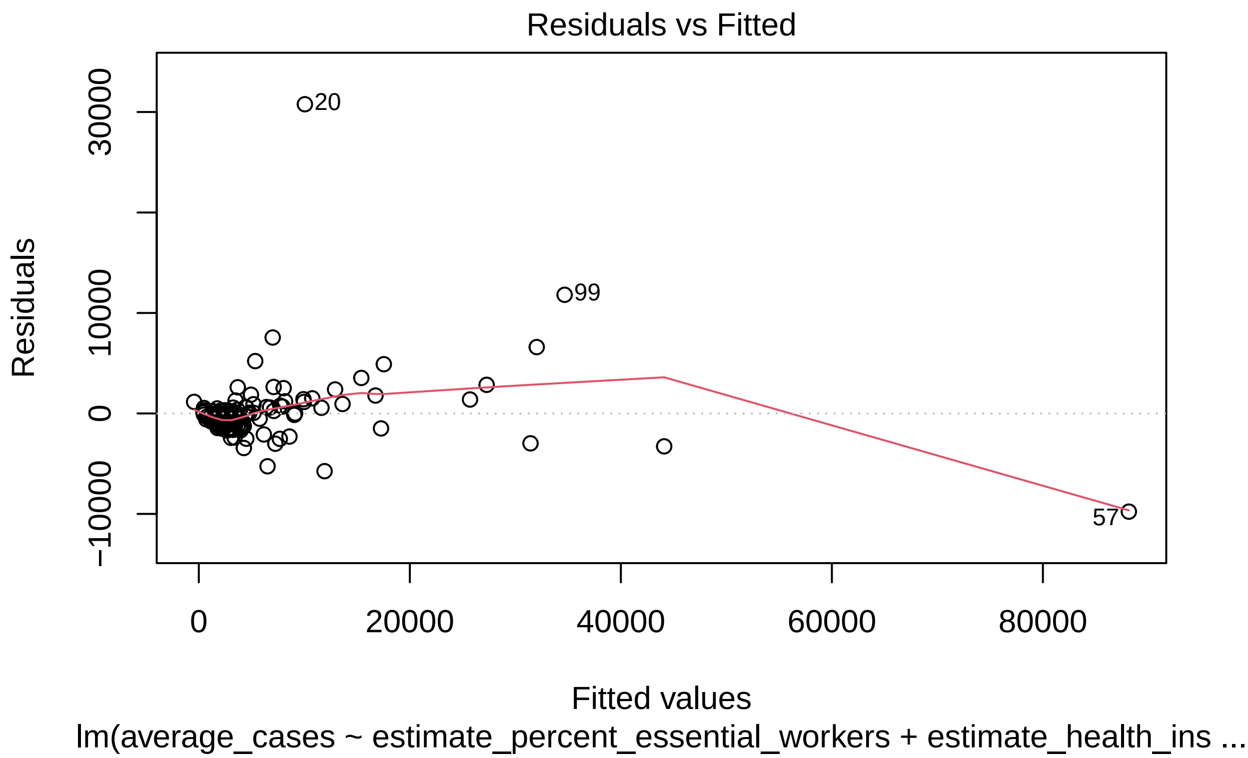

Residuals vs Fitted
Residuals
Fitted values
lm(average_deaths ~ estimate_percent_essential_workers + estimate_health_in ...
600
400
200
0
−200
0
200
400
600
800
1000
1200
20
70
37

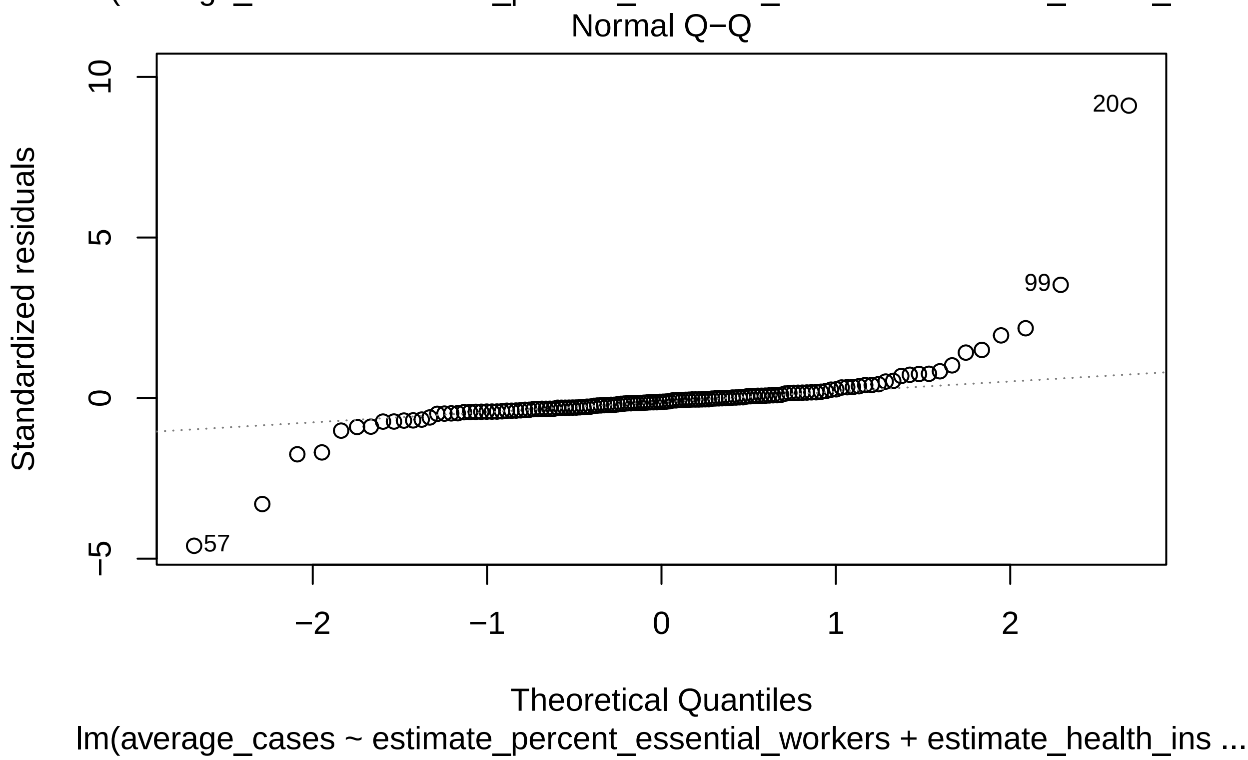

Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(average_deaths ~ estimate_percent_essential_workers + estimate_health_in ...
10
8
6
4
2
0
−2
−2
−1
0
1
2
20
70
37

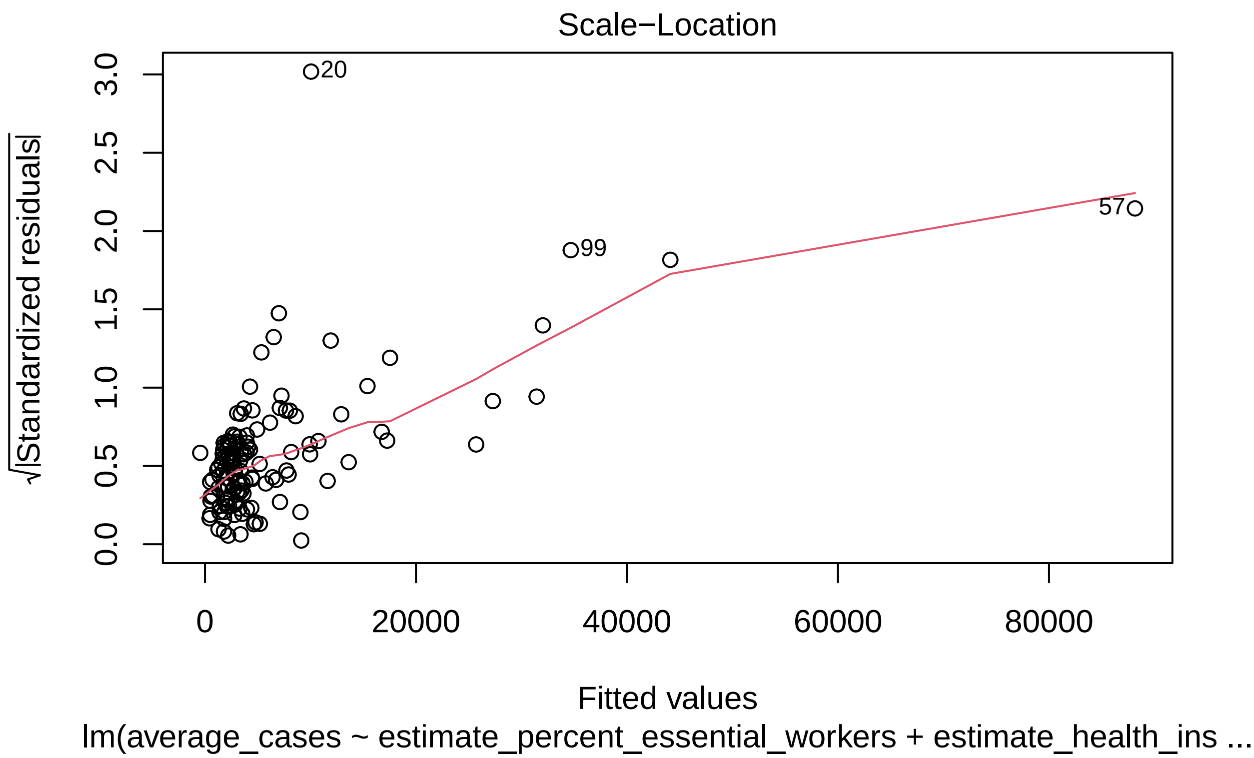


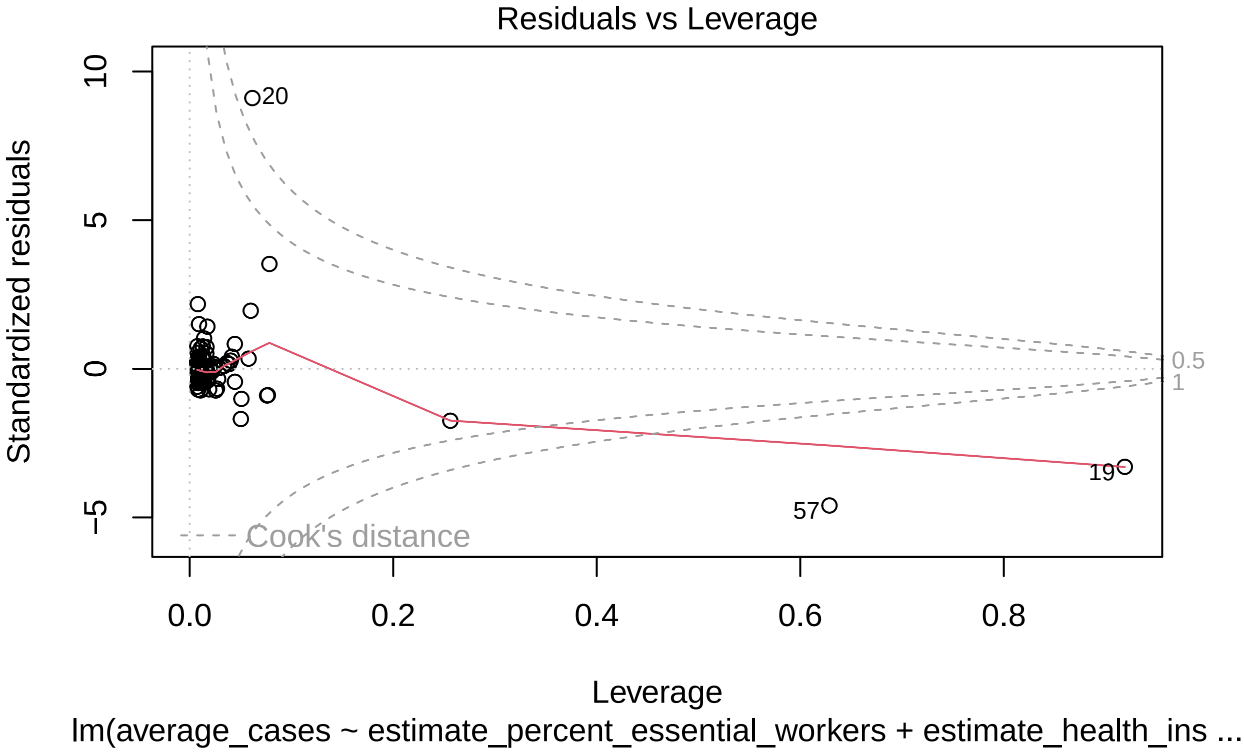



```
plot(average_cases_reduced)
```


Residuals vs Fitted
Residuals
30000
10000
0
−10000
20
99
57
0
20000
40000
60000
80000
Fitted values
lm(average_cases ~ estimate_percent_essential_workers + estimate_health_ins ...


Normal Q−Q
Standardized residuals
10
5
0
−5
20
99
57
−2
−1
0
1
2
Theoretical Quantiles
lm(average_cases ~ estimate_percent_essential_workers + estimate_health_ins ...


Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
20000
40000
60000
80000
20
99
57
Fitted values
lm(average_cases ~ estimate_percent_essential_workers + estimate_health_ins ...


Residuals vs Leverage
Standardized residuals
−5
0
5
10
0.0
0.2
0.4
0.6
0.8
20
57
19
0.5
1
Cook's distance
Leverage
lm(average_cases ~ estimate_percent_essential_workers + estimate_health_ins ...