Age Distribution vs COVID-19 Case Count

Mrinal Chawla, Roy Katende, Yusuf Liu
Group 4 - W203 Fall 2020

Introduction

Problem: The impact of age distribution within each state of the USA on the case rate of COVID-19.

Solution:A descriptive model that can be a useful for each state to have a baseline on how COVID-19 affected the population, and where to focus recovery efforts

The Data

- Age Distribution
- Case Rate / 100000
- Population density / square mile
- Tests per 100K

Age Distribution

-- measured in 6 groups as a percentage (0-18, 19-25, 26-34, 34-54, 55-64, 65+)
-- provides window into segmentation of population, and how case rates are affected

Case Rate / 100000

-- measured as decimal number
-- normalized way to compare states

Population Density

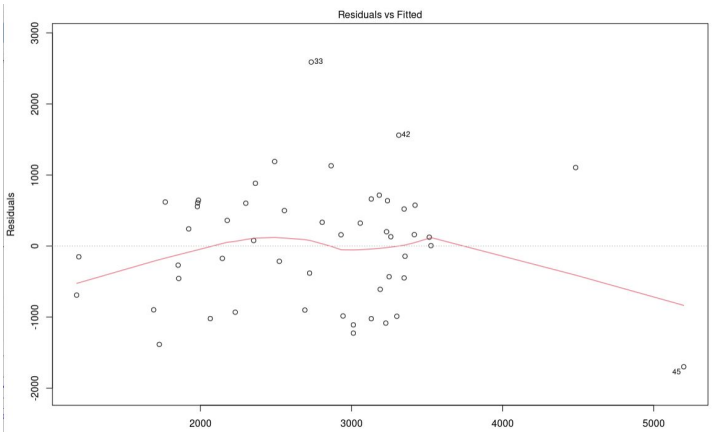
-- measured as rate per square mile
-- provides insight into transmission

Tests per 100K

-- measures as decimal number
-- needed to control for testing rate, as case rate will not be affected without tests

Model 1

- Uses only age distribution
- $R^2 = 0.4512$
- F-test
  - P-value: 0.0001
  - 44 DF
- Only 19-25 significant



Outliers:

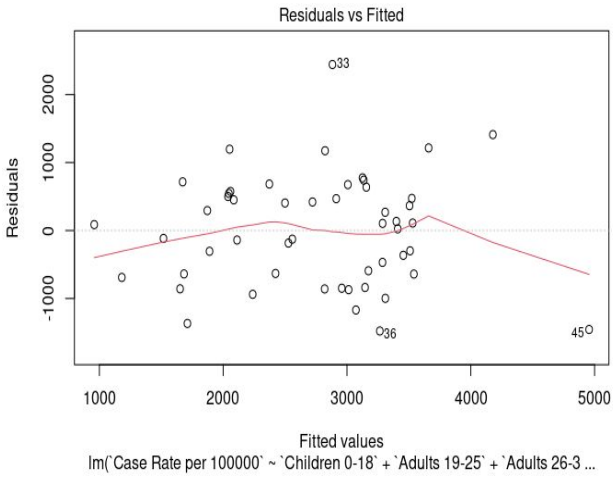
33: New Mexico

42: South Carolina

45: New York

Model 2

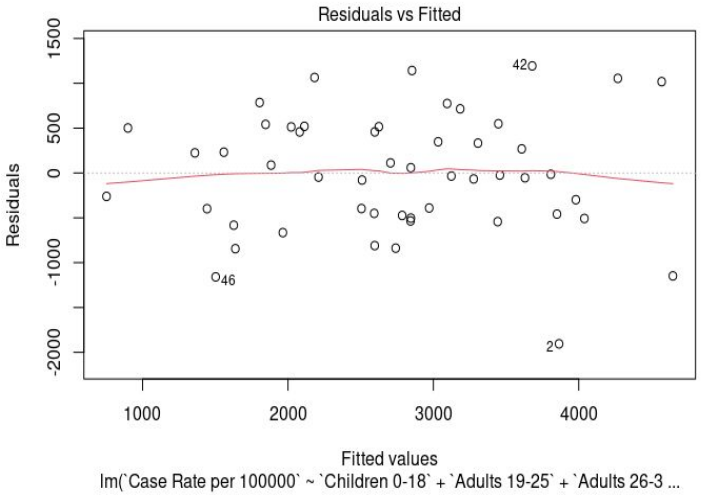
- Added in population density per square miles, Nonelderly Adults Who Have A Pre-Existing, Condition and the transformed variable Closed Business Days.
- $R^2 = 0.4865$
- F-test
  - 41 DF and P-value: 0.0005
- Only 19-25 very significant, and 0-18 less significant.



For the second model, we added in a few new features to help the model describe the problem better, log of population density per square miles, nonelderly adults who have a pre-existing condition , and the transformed variable closed business days. The two significant variables appear to be Children 0-18 and Adults 19-25. As we can see with the new features being added, the r2 slight improved to 0.49. The outlier is 33 New Mexico,45 new york. This model suggests that the percentage of population in the two lowest age categories are the best indicators for case rate in a state.

Model 3

- Tests per 100K
- Mandate face mask use
- $R^2 = 0.6631$
- F-test
  - 39 DF and P-value: 2.366e-06
- Significant Variables:
  - Children 0-18
  - Pop density / sq. mile
  - 65+



- 2 variables are added “Tests per 100K” and “Mandate Face Mask Use” (Binary Y/N) 60/40
- The addition of these makes 3 older variables significant (the outer edges of the age distribution)
- The Adjusted R2 in Detailed Model reduces from 11 to 9 because of the significant variables
- Practical Significance is Mask usage & tests affect the case rate
- Residual vs Fitted shows that it is becoming more of a linear model with these added variables

Thank You

In conclusion, our final model with age distribution, and policies, can serve as an important tool for pandemic mitigation and
Recovery. Based on the feature importance, we found that younger generation have a huge impact on the case rate, especially the 0-18 and 19-25 age group. Thus, the model will help the government highlight which state populations are at risk. By knowing how the population was affected,
governments can make plans and policies to help the state recover.
Thank you for listening.