Midertm Project

Anna Cook

12/10/2020

Abstract

The purpose of this analysis is to explore the relationships between race/ethnicity, poverty rate, and voting tendencies across states in the U.S. A multilevel linear model was fit to the data with random intercepts for each state. The results show _____. While there are important limitations to consider, these results suggest that...

Introduction

Amid the Covid-19 pandemic as well as Donald Trump's presidency, the issues of wealth and racial disparities in the United States have become more exposed, and appear to be more prevalent now than ever before. The analysis presented here is aimed at better understanding how these issues are related to one another, by examining poverty rates for different racial/ethnic groups across states in the U.S., as well as presidential election results for those states. The research question is two-fold: First, I am interested in determining whether there is a relationship between poverty rate and racial/ethnic groups across states. For example, are there racial groups who are economically disadvantaged in some states/regions but not in others? Second, I am interested in determining whether the poverty rate and racial/ethnic groups in various states is associated with the voting behavior in those states. The data was collected from two different sources. The first set of data was collected from _____ and contains information on the poverty rate for various racial/ethnic groups, organized by U.S. state and year (2009-2017). Although the original dataset included ___ racial/ethic groups, I chose to focus my analyses on only ___ of those: black, hispanic, white, asian, and **. The second set of data was collected from** _ and contains information on U.S. presidental election results, organized by state and election year. The original dataset includes elections dating back to _____, but I filtered the dataset to include only elections between 2008-2016 since those years are most closely aligned with the poverty dataset.

Modeling

To examine an association between race/ethnicity and poverty rate for different states, I fit a multilevel linear model with intercepts varying for each state, using the rstanarm package. Next, to examine an association between voting behavior and race/ethnicity and poverty rate, I fit a multilevel linear model with intercepts varying for each state.

In order to validate these models, I used a posterior predictive check to compare the data's distribution to that of predictive simulations, in addition to examining the residual plot for each model. The results of these validation methods are described below.

Results

Discussion

One major limitation to this study is that the sample sizes are too small. Because there is only one poverty rate per racial/ethnic group per state per year, there isn't enough data to fit reliable models when using all of those variables as predictors. This probably may be partially alleviated by collapsing across states, but this

wouldn't allow us to see how patterns vary by state. Another limitation is that while I am interested in voting behavior, my dataset only contained information on presidential elections which are held once every four years. This also leads to too small of sample sizes since the datasets only go back to 2008. The analyses may have been more reliable with more information; for example, if the dataset had contained voting behavior for more years, or on local elections as well as presidential elections. Both of these limitation led to poor performance of any of the models I tried fitting. Using different sets of predictors, collapsing across year, and using different types of models only led to very small changes, and in the end, no model fit the data very well. This was an important learning experience for me, as I now have a better sense of what to look for when I am searching for sufficient data to answer my research questions in the future.

Appendix

References

Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01.

Goodrich B, Gabry J, Ali I & Brilleman S. (2020). rstanarm: Bayesian applied regression modeling via Stan. R package version 2.21.1 https://mc-stan.org/rstanarm.

H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

Hadley Wickham, Jim Hester and Romain Francois (2018). readr: Read Rectangular Text Data. R package version 1.3.1. https://CRAN.R-project.org/package=readr

Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2020). dplyr: A Grammar of Data Manipulation. R package version 1.0.2. https://CRAN.R-project.org/package=dplyr

Hadley Wickham (2020). tidyr: Tidy Messy Data. R package version 1.1.2. https://CRAN.R-project.org/package=tidyr

Stefan Milton Bache and Hadley Wickham (2014). magrittr: A Forward-Pipe Operator for R. R package version 1.5. https://CRAN.R-project.org/package=magrittr