If Income Influence Vote Preference

Zizhuo Huang

2020/12/20

#Abstract

In problem set 3, we use MRP to model the proportion of voters who will vote for Joe Biden. In this report, I use the same dataset to investigate if there is a relationship between income and vote preference.

#Keywords

Propensity Score Matching, Logistic Regression, Causual Inference, Income, Vote Preference

#Introduction

2020 American Federal Election is the most concrened issue among the whole world. When the United States was just founded, the founders of the United States designed the electoral college system. At that time, the main purpose was to prevent politicians from making false promises to the voters to cheat votes, and the “electors” would indirectly choose the president, so as to avoid disadvantages. But that meaning is gone. The main purpose of this system is to respect the rights of states, which is the embodiment of American democracy’s decentralization and respect for the rights of local states.

For predicting the election result and study what factors affect the citizens’ voting preference, I use a data from American Community Surveys. Although there are many other suvjective problems which impact the vote for election. The topic I will mostly diiscuss in this reprot is that if the amounts of income influence the final vote preference.

To do this, I use the recent learning knowledge which are propensity score matching combining with causual inference to complete my report. Firstly, I describe the simulation study, the data, and the model that was used to perform the propensity score analysis. Results of the propensity score analysis are provided in the Results section. Finally, I discuss the weakness and the next step we can do to make the guess more credible.

#Methodology

For data, first step is to read in the raw data and add the labels to all variables. Then keeping some variables to make sure it exists in census data as well except for vote intention. I choose some variables, registration, age, gender, education, state and household income and adjust data types. Next step is to filter on survey data, only leaving the people that are both registered & intented to vote (Assuming people will vote unless they explicitly say no). Then I create a new variable, income_high, which is those who have over than $150,000 household income and adjust the new dataset, data3.

For model, at first, I calculate the propensity score and use the forecast to create matches, using a matching function from the arm package. This finds which is the closest of the ones that were not treated, to each one that was treated. Then I reduce the dataset to just those that are matched. There were 389 treated, so we expect a dataset of 778 observations.At last, I examine the ‘effect’ of being treated on average spend in the ‘usual’ way. Besides, it is logistic to predict “voting for Biden”. The last step is to create two logistic regression model to complete the investigation. The treatment is income_high, outcome is vote_202 and the instrument variable is age. Here are the reults.

```
##
## Call:  glm(formula = vote_2020 ~ age, family = "binomial", data = data4)
##
## Coefficients:
## (Intercept)          age
##     -0.6054       0.0074
##
## Degrees of Freedom: 777 Total (i.e. Null);  776 Residual
## Null Deviance:       1065
## Residual Deviance: 1063  AIC: 1067
```

```
##
## Call:  glm(formula = income_high ~ age, family = "binomial", data = data4)
##
## Coefficients:
## (Intercept)          age
##    0.056012    -0.001215
##
## Degrees of Freedom: 777 Total (i.e. Null);  776 Residual
## Null Deviance:       1079
## Residual Deviance: 1078  AIC: 1082
```

#Results

For the regression model of vote preference and age, gender, education, state and income_high. The Intercept is 0.081 so when everthing else is zero, the probablity of voting for Biden is 0.081. I find that gender and income are two significant factors to influence the vote. The coefficient are -0.664 when male and -0.486 when high separately. That is to say, when eveything else stay same. If the voter is male or have income over $150,000, the probabilty they vote for Biden will decrease by 0.664 and 0.486 resprctively.

Combining two logistic models which are age versus vote and income_high versus vote, I find that the coeficients are 0.0074 and -0.001215 separately. So the age and income_high has negative correlation while age and vote preference have positive correlation. Additionally, income_high and vote for Biden have negative correlation, which is consistent with the former result.

#Disccussion

In conclusion, after doing propensity score matching and causual inference, the amounts of income truly affect voters’ vote preference and they have the negative correlation. That is to say, when voter have high income, they will not vote for Biden and they tend to choose Trump to be the president.

However, the model still have some weakness like I just choose a few variavles to do the research and the propensity score just matching the similar observations which cannot be exactly same.

#Reference

Tausanovitch, Chris and Lynn Vavreck. 2020. Democracy Fund + UCLA Nationscape, October 10-17, 2019 (version 20200814). Retrieved from [URL]. Steven Ruggles, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas and Matthew Sobek. IPUMS USA: Version 10.0 [dataset]. Minneapolis, MN: IPUMS, 2020. https://doi.org/10.18128/D010.V10.0 (https://doi.org/10.18128/D010.V10.0) Wei Wang, David Rothschild, Sharad Goel, Andrew Gelmana, chttps://www.microsoft.com/enus/ research/wp-content/uploads/2016/04/forecasting-with-nonrepresentative-polls.pdf, 2014