

Statistical Inference on Factors Affecting Voting Results of the 2020 U.S. Election

Yicheng Ding

2020/12/21

Introduction

When it was announced that Joe Biden was elected President of the United States, the 2020 U.S. election came to an end. Under the influence of COVID-19, this year's election is very difficult and dramatic. In Georgia, thousands of voters waited hours just for voting during early voting. Since the pandemic, surge mail-in ballots make counting votes extremely difficult and make the election period stretched. In rural parts of America, voters may have to drive for hours to reach an in-person polling station (Robin, para. 8-15).

In this report, the purpose is investigating the effect of three variables (region, race, and education) to the voting result of 2020 U.S. election. Thus, the null hypothesis of this report is that there is no relationship between region, race, and education and the voting result. If the results show significant relationship between our dependent variable and independent variables, this will provide new campaign ideas for candidates in the 2024 U.S. election.

One data sets will be used to investigate the effect of three variables (region, race, and education) to the voting result of 2020 U.S. election. In the Methodology section, the data and the model will be described. The results of the relationship between region, race, and education and the voting result will be provided in the Results section. Inferences of this data with conclusion, weakness and next step will be in Conclusion section (Samantha-Jo, para. 8).

Methodology

Data

In this report, one data sets will be used which is individual-level public opinion survey data from the Democracy Fund + UCLA Nationscape.

Table 1: Voting result with census_region, race_ethnicity and education

```
## # A tibble: 4,868 x 5
##   vote_2020 census_region race_ethnicity education vote_result
##   <chr>      <chr>        <chr>      <chr>      <dbl>
## 1 Donald Tru~ Midwest      White      Associate's degree      0
## 2 Donald Tru~ South        White      Bachelor's degree      0
## 3 Donald Tru~ South        White      High school diplo~     0
## 4 Donald Tru~ West         White      High school diplo~     0
## 5 Joe Biden  Northeast    White      College, no degree     1
## 6 Joe Biden  South        Black or African Am~ College, no degree     1
## 7 Donald Tru~ West         White      College, no degree     0
## 8 Donald Tru~ South        White      Bachelor's degree      0
```

```
## 9 Donald Tru~ South      White      Bachelor's degree      0
## 10 Joe Biden  South      White      College, no degree      1
## # ... with 4,858 more rows
```

In table 1, we can find out the voting result of 2020 U.S. Election. Our dependent variable is vote result. 0 is stand for vote to Donald Trump and 1 is stand for vote to Joe Biden. Census region, race ethnicity and education are all the independent variables that we need to study.

Method

Logistic regression will be used in this analysis because it is the appropriate regression when the dependent variable is binary.

#References

Robin, Levinson-King. (October 20th, 2020). US election 2020: Why it can be hard to vote in the US. BBC News. Retrieved from <https://www.bbc.com/news/election-us-2020-54240651> Samantha-Jo, Caetano. (December 2020). Final Project – Additional Instructions.