

Forecasting 2020 US Election Outcomes: The Role of Employment, Education, gender, and Race*

Sean Liu

March 16, 2024

hi Second sentence. Third sentence. Fourth sentence.

*Code and data are available at: <https://github.com/Shuuu/Politics.git>

1 Introduction

Back in 2020, the US was having an important presidential election contest between the Democratic party and the Republican party. The Democratic ticket of former vice president Joe Biden and the Republican president, Donald Trump. The election saw the highest voter turnout by percentage since 1900, with each of the two main tickets receiving more than 74 million votes. In particular we are interested in whether we can forecast who a respondent is likely to vote for, based on knowing their employment status, highest level of education, and race. That means we are interested in a data-set with variables for who an individual voted for, and some of their characteristics, such as employment status, education level, gender, and their race. The data is from the 2020 election data set from Cooperative Election Study (CES) (Schaffner, Ansolabehere, and Luks 2021).

In this study, we used a logistic regression model to predict the 2020 election results, utilizing data from the Cooperative Election Study (CES). This model is well-suited for binary outcomes, such as predicting whether an individual voted for Trump or Biden. Our analysis aims to assess the probability of victory for each candidate, considering a variety of factors including employment status, education level, gender, and their race. The primary objective is to determine the actual support levels for Trump and Biden.

The remainder of this paper is structured into different sections. Section 2 shows the data that are used for our study. It includes some graphs to demonstrate different groups of respondents in our data. Section 3 builds the model and discusses its justification and explanation. **?@sec-result** highlights the results of the predictions using tables and graphs. **?@sec-discussion** contains discussions that conducted based on the findings, which addresses the voting prediction results based on race, region, employment status, and the influence of COVID-19 and in-mail voting systems.

2 Data

2.1 Source of Data

Our primary data is from the 2020 Cooperative Election Study, CES Schaffner, Ansolabehere, and Luks (2021). The data includes a nationally representative sample of 61,000 American adults. Schaffner, Ansolabehere, and Luks (2021) includes the data from the survey, a full guide to the data, and the questionnaires.

?@fig-education shows the relationship between respondents' educational level, gender and their voting preferences. In both gender, respondents with higher educational level tends to lean more towards Biden.

3 Model

3.1 Model set-up

The model that we are interested in is:

[

$$y_i | \pi_i \sim \text{Bern}(\pi_i) \quad (1)$$

$$\text{logit}(\pi_i) = \alpha + \beta_1 \times \text{employment_status}_i + \beta_2 \times \text{race}_i + \beta_3 \times \text{education}_i + \beta_4 \times \text{gender}_i \quad (2)$$

$$\alpha \sim \text{Normal}(0, 2.5) \quad (3)$$

$$\beta_1 \sim \text{Normal}(0, 2.5) \quad (4)$$

$$\beta_2 \sim \text{Normal}(0, 2.5) \quad (5)$$

$$\beta_3 \sim \text{Normal}(0, 2.5) \quad (6)$$

$$\beta_4 \sim \text{Normal}(0, 2.5) \quad (7)$$

]

Where y_i is the political preference of the respondent, it equals to 1 if the respondent voted for Biden and 0 for Trump, $\text{employment_status}_i$ is the employment status of the respondent, race_i is the race of the respondent, education_i is the highest educational level of the respondent, and lastly gender_i is the gender of the respondent. The coefficients $(\beta_1, \beta_2, \beta_3)$ are being assumed as a normal distribution with a mean of 0 and a standard deviation of 2.5 for each parameter. Lastly, p_i is the probability of voting for Biden. We run the model in R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2022).

3.2 Model justification

We anticipate a positive correlation between individuals identifying as Black, Asian, or Hispanic and their support for Biden. As Biden's proposal to more advanced racial equity in the US (Sprunt 2020). Most traditional white voters are more likely to support Trump. As Trump's cultural and the atmosphere of Make America great again. Which tend to these group of people to support him (Gene Demby 2020).

For the education level, we expect respondents who has a higher educational level to be leaning towards to support Biden. As people with higher educational levels are more open to diverse perspectives and the new directions in education that Biden proposed (Wong 2020).

Lastly the voting preferences for employed voters and the difference in gender are harder to determine.

Gene Demby, Shereen Marisol Meraji. 2020. "The White Elephants in the Room." *Npr*.
<https://www.npr.org/2020/11/17/935910276/the-white-elephants-in-the-room>.

- Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. “Rstanarm: Bayesian Applied Regression Modeling via Stan.” <https://mc-stan.org/rstanarm/>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Schaffner, Brian, Stephen Ansolabehere, and Sam Luks. 2021. “Cooperative Election Study Common Content, 2020.” Harvard Dataverse. <https://doi.org/10.7910/DVN/E9N6PH>.
- Sprunt, Barbara. 2020. “Biden Details Proposal to Advance Racial Equity in America.” *Npr*. <https://www.npr.org/2020/07/28/896132054/biden-details-proposal-to-advance-racial-equity-in-america>.
- Wong, Kenneth K. 2020. “The Biden Presidency and a New Direction in Education Policy.” *Brookings*. <https://www.brookings.edu/articles/the-biden-presidency-and-a-new-direction-in-education-policy/>.