Analyzing Factors Related to Voter Preferences in the United States

Ravnit Lotay

March 16, 2024

This study investigates the factors influencing voter support in the 2022 US presidential election, utilizing logistic regression on CCES 2022 data. The results of the analysis will be crucial in understanding why individuals or demographics vote for a certain individual or party,

Table of contents

1 Introduction

The goal of the Cooperative Election Study (CES) is to investigate how Americans vote and perceive their electoral experiences, how they hold their representatives accountable in elections, and how their behavior and experiences vary. The study used a sample of 60,000, allowing it to account for variations across different legislative districts.

It is important to note that the 2022 CES data has adjusted the input data via a weightening process to ensure samples are valid and representative.

2 Data

The Dataset contains 60,000 cases in which individuals responsed to survey questions including, but not limited to: voting prefrence, age, gender and highest level of education. The mentioned factors will be used to draw inferences about voter preferences via a logistic regression model.

2.1 Data Visualizations

Within the CES data, there are relations between the different variables. Below are two figures showing how gender, age, and education relate to voter preference.

3 Model

To assess the roles of gender, education, and age on voting preferences, a logistic regression model is used.

3.1 Model Creation

To create the model, let y_i represent the respondent's political preference, defined as 1 for Trump and 0 for Biden. Gender is denoted by g_i, and age by a_i.

The logistic regression model can be formed as follows:

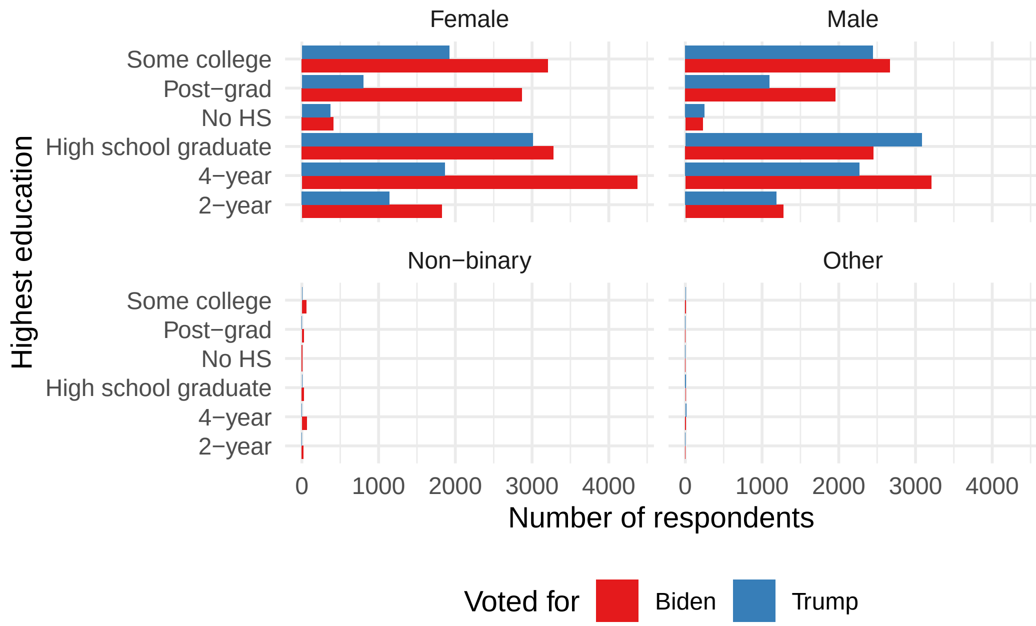

Figure 1: Distribution of Presidential Preferences by gender and education

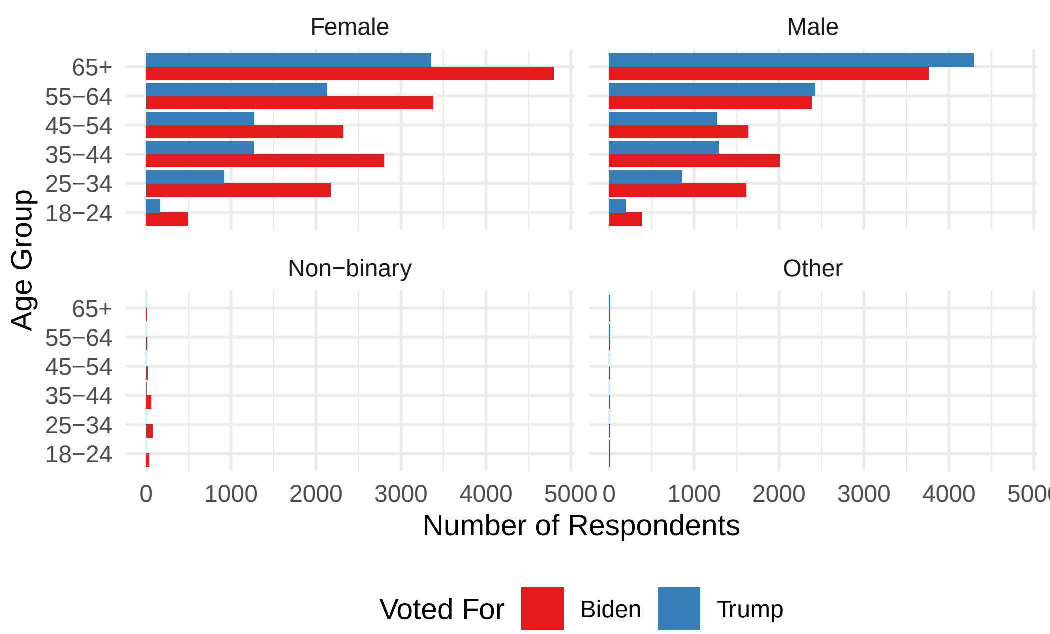

Figure 2: Distribution of Presidential Preferences by age and education

$$
\begin{aligned}
y_i|\pi_i &\sim \text{Bernoulli}(\pi_i) \\
\text{logit}(\pi_i) &= \beta_0 + \beta_1 \times g_i + \beta_2 \times e_i \\
\beta_0 &\sim \text{Normal}(0, 2.5) \\
\beta_1 &\sim \text{Normal}(0, 2.5) \\
\beta_2 &\sim \text{Normal}(0, 2.5)
\end{aligned}
$$

Where: - y_i is the political preference of the ith respondent. - π_i is the probability of supporting Trump for the ith respondent. - g_i represents the gender of the ith respondent. - e_i represents the education level of the ith respondent. - β_0, β_1, and β_2 are the parameters of the model, representing the intercept, the effect of gender, and the effect of education on the likelyhood of supporting Trump, respectively.

4 Results

The logistic regression model provides invaluable information related to voter preferences.

4.1 Results Table

Below, the results of the logistic regression model are shown in a statistical form.

In the next section, the results will be interpreted in a simpler manner.

5 Discussion

5.1 Model Interpreted Results

The summary of the results, as shown above, indicate the following:

1. Females and those identifying as "Non-binary" are more strongly against voting for Trump when compared to the reference group.
2. Those identifying as "Other" are more likely to vote for Trump than the reference group.
3. Those that have completed a 4-year college or university program are much less likely to vote for Trump than those with no highschool education or those with a highschool diploma.

Table 1: Results of logistic regression on the voted for Trump

	Support Trump
(Intercept)	−1.603
	(0.045)
genderFemale	−0.427
	(0.020)
genderNon-binary	−2.274
	(0.276)
genderOther	0.226
	(0.268)
educationSome college	0.665
	(0.034)
education4-year	0.387
	(0.034)
educationHigh school graduate	1.008
	(0.034)
education2-year	0.693
	(0.039)
educationNo HS	1.011
	(0.064)
age	0.015
	(0.001)
Num.Obs.	47 466
R2	0.053
Log.Lik.	−30 838.276
ELPD	−30 848.3
ELPD s.e.	61.4
LOOIC	61 696.6
LOOIC s.e.	122.8
WAIC	61 696.6
RMSE	0.48

The coefficient for woman, is seen as -0.427, while the coefficient for Non-binary is -2.274. This, in essence, implies that these groups are much more opposed to voting for Trump than our baseline reference group of males.

As for age, due to the positive coefficient, it is evident that the older an individual is, the more likely they are to vote for Trump.

It is important to note that since these values are based on a reference group (males for the gender attribute), a positive coefficient does not inherently mean that the majority of the group is voting for Trump. To further prove this statement, it is clear that despite many coefficients being positive in the above table, Donald Trump was not elected in the last presidential election.

5.2 Weaknesses

Considering the data collected and analyzed, there are several weaknesses that could be addressed. First, there are relatively few Non-binary and Other values in the gender column, and as such they could be grouped together or otherwise adjusted for. This is necessary as logistic regression is typically influenced greatly by outliers. Second, additional parameters, such as race, could be used to improve the quality of the model.

6 References

Arel-Bundock, Vincent. 2021. "Modelsummary: Summary Tables and Plots for Statistical Models and Data: Beautiful, Customizable, and Publication-Ready." *Journal of Open Source Software* 6 (61): 3167. https://doi.org/10.21105/joss.03167.

Auguie, Baptiste. 2017. *gridExtra: Miscellaneous Functions for "Grid" Graphics*. https://CRAN.R-project.org/package=gridExtra.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. *Rstanarm: Bayesian Applied Regression Modeling via Stan*. https://mc-stan.org/rstanarm.

Leeper, Thomas J. 2021. *Dataverse: An r Client for Dataverse 4 Repositories*. https://github.com/IQSS/dataverse-client-r.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, et al. 2019. *Tidyverse: Easily Install and Load the Tidyverse*. https://tidyverse.tidyverse.org.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2021. *Dplyr: A Grammar of Data Manipulation*. https://dplyr.tidyverse.org.

Wickham, Hadley, and Jim Hester. 2021. *Readr: Read Rectangular Text Data*. https://readr.tidyverse.org.

Wickham, Hadley, and Evan Miller. 2021. *Haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files*. https://CRAN.R-project.org/package=haven.

Xie, Yihui. 2021. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.