# 2022 US Political Survey Analysis*

Yimiao Yuan

March 19, 2024

## 1 Introduction

In this paper, I use the 2022 Cooperative Election Study (CES) (Schaffner, Ansolabehere, and Shih 2023) as our primary dataset. This is a long-established annual political survey of US. In 2022, there were 61,000 American adults completed the survey. Since it is difficult to observe such a large dataset, this paper will only analyze respondents who are registered to vote and focus on Biden and Trump's results.

The dataset is downloaded, cleaned and analyzed using statistical programming language R (R Core Team 2023), with packages `tidyverse` (Wickham et al. 2019), `here` (Müller 2020), `rstanarm` (Brilleman et al. 2018), `modelsummary` (Arel-Bundock 2022), `knitr` (Xie 2014), `marginaleffects` (Arel-Bundock 2024).

## 2 Model

The models are created to analyze the relationship between voter preferences and voter's race. In the model, $y_i$ represents whether an individual voted Biden or Trump. For all parameters, I use a normal distribution with a mean of 0 and a standard deviation of 2.5.

*Code and data are available at: https://github.com/YimiaoYuan09/US_Election_2022

Table 1: Logistic and Poisson model results

| | Logistic | Poisson |
|---|---|---|
| (Intercept) | 1.718 | −0.182 |
| raceBlack | −0.011 | 0.008 |
| raceHispanic | −0.848 | −0.173 |
| raceMiddle Eastern | −1.103 | −0.283 |
| raceNative American | −1.145 | −0.314 |
| raceOther | −0.871 | −0.424 |
| raceTwo or more races | −1.037 | −0.264 |
| raceWhite | −1.710 | −0.506 |
| Num.Obs. | 1000 | 1000 |
| R2 | 0.071 | |
| Log.Lik. | −644.354 | −882.635 |
| ELPD | −654.7 | −888.2 |
| ELPD s.e. | 9.8 | 9.7 |
| LOOIC | 1309.3 | 1776.5 |
| LOOIC s.e. | 19.5 | 19.4 |
| WAIC | 1307.4 | 1776.3 |
| RMSE | 0.48 | 0.48 |

### 2.1 Logistic Regression

$$
\begin{align}
y_i|\pi_i &\sim \text{Bern}(\pi_i) \tag{1}\\
\text{logit}(\pi_i) &= \alpha + \beta_1 \times \text{race}_i \tag{2}\\
\alpha &\sim \text{Normal}(0, 2.5) \tag{3}\\
\beta_1 &\sim \text{Normal}(0, 2.5) \tag{4}\\
\beta_2 &\sim \text{Normal}(0, 2.5) \tag{5}
\end{align}
$$

### 2.2 Poisson Regression

$$
\begin{align}
y_i|\lambda_i &\sim \text{Poisson}(\lambda_i) \tag{6}\\
\text{logit}(\lambda_i) &= \alpha + \beta_1 \times \text{race}_i \tag{7}\\
& \tag{8}
\end{align}
$$

Table 2: Negative Binomial model results

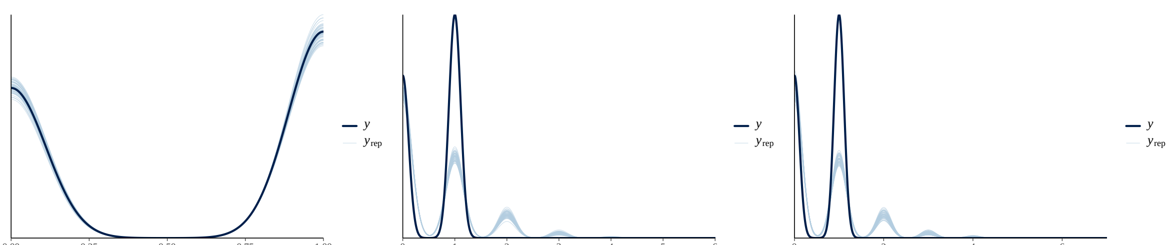
| | mean | mcse | sd |
|---|---|---|---|
| (Intercept) | -0.1878955 | 0.0063387 | 0.2270506 |
| raceBlack | 0.0149745 | 0.0067156 | 0.2470082 |
| raceHispanic | -0.1686325 | 0.0066472 | 0.2598085 |
| raceMiddle Eastern | -0.2938729 | 0.0089301 | 0.4008122 |
| raceNative American | -0.3299962 | 0.0092645 | 0.4503338 |
| raceOther | -0.4527730 | 0.0142343 | 0.7959042 |
| raceTwo or more races | -0.2753266 | 0.0090451 | 0.4122810 |
| raceWhite | -0.4969023 | 0.0064838 | 0.2344021 |
| reciprocal_dispersion | 13.6220229 | 0.0396286 | 2.6448908 |
| mean_PPD | 0.5797825 | 0.0005669 | 0.0362918 |
| log-posterior | -919.0171027 | 0.0537600 | 2.1960227 |

### 2.3 Model Results

The results are summarized in Table 1 and Table 2. Poisson and Negative Binomial model have similar results, while Logistic model has a slightly different one. Logistic model has an intercept of around 1.7, but poisson and negative binomial model is about -0.2. Other parameters have similar trends in these three models, but with different values.

### 2.4 Posterior predictive check

Figure 1 shows the posterior distribution for three models. According to the graphs, I can conclude that Logistic Regression is a better choice for current situation. Also, since our outcome (vote for Biden/Trump) is a binary variable, Logistic regression fits better. Poisson regression and negative binomial regression are used where the dependent variable represents the count of events occurring at a fixed time or spatial interval.




(a) Logistic Regression (b) Poisson Regression (c) Negative Binomial Regression

Figure 1: Posterior distribution for models

## References


Arel-Bundock, Vincent. 2022. "modelsummary: Data and Model Summaries in R." *Journal of Statistical Software* 103 (1): 1–23. https://doi.org/10.18637/jss.v103.i01.

———. 2024. *Marginaleffects: Predictions, Comparisons, Slopes, Marginal Means, and Hypothesis Tests*. https://CRAN.R-project.org/package=marginaleffects.

Brilleman, SL, MJ Crowther, M Moreno-Betancur, J Buros Novik, and R Wolfe. 2018. "Joint Longitudinal and Time-to-Event Models via Stan." https://github.com/stan-dev/stancon_talks/.

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

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 Marissa Shih. 2023. "Cooperative Election Study Common Content, 2022." Harvard Dataverse. https://doi.org/doi.org/10.7910/DVN/PR4L8P.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC.