My title*

My subtitle if needed

First author Another author

March 12, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

1 Introduction

2 Data

3 Model

3.1 Model set-up

The goal of our modeling strategy is to forecast if a person voted for Biden, based only on knowing their gender, race, and gun ownership status.

The model that we are interested in is:

$$
\begin{align}
y_i|\pi_i &\sim \text{Bern}(\pi_i) \tag{1}\\
\text{logit}(\pi_i) &= \alpha + \beta \times \text{gender}_i + \gamma \times \text{education}_i + \delta \times \text{gun}_i \tag{2}\\
\alpha &\sim \text{Normal}(0, 2.5) \tag{3}\\
\beta &\sim \text{Normal}(0, 2.5) \tag{4}\\
\gamma &\sim \text{Normal}(0, 2.5) \tag{5}\\
\delta &\sim \text{Normal}(0, 2.5) \tag{6}
\end{align}
$$

Where:

*Code and data are available at: LINK.

- y_i is the binary outcome variable, representing who respondent i voted for and equal to 1 if Biden and 0 if Trump,
- π_i is the probability that respondent i voted for Biden,
- $gender_i$ is a predictor variable, representing the gender of respondent i,
- $race_i$ is a predictor variable, representing the race of respondent i, and
- gun_i is a predictor variable, representing the gun ownership status of respondent i.

We used a logistic regression model in a Bayesian framework using the package `rstanarm` (Goodrich et al. 2022), which we will briefly describe here. Logistic regression is a type of generalized linear model. It is a tool for data exploration and used when we are interested in the relationship between a binary outcome variable and some predictor variables.

The foundation of logistic regression is the Bernoulli distribution and logit function. The Bernoulli distribution is a discrete probability distribution having two possible outcomes, "1" and "0", in which "1" occurs with probability p and "0" occurs with probability $1-p$. Logistic regression is still a linear model, because the predictor variables enter in a linear fashion (Wickham et al. 2019). Hence, the logit function links the Bernoulli distribution to the machinery we use in linear models (Wickham et al. 2019).

In our model, we also have the parameters α, β, γ, and δ in addition to the variables. The parameter α is the intercept and the parameters, β, γ, and δ, are the slope coefficients. We specify prior probability distributions for each of the parameters in our model, but these are just the default priors that `rstanarm` (Goodrich et al. 2022) uses (Normal distribution with mean and standard deviation of 0 and 2.5).

3.2 Model justification

4 Results

5 Discussion

5.1 First discussion point

5.2 Second discussion point

5.3 Third discussion point

5.4 Weaknesses and next steps

Appendix

A Additional data details

B Model details

References

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

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.