Election Preferences in 2022 based on Gender and Education*

Rahul Gopeesingh

March 16, 2024

There are many factors that can influence someone's voting preference. Two of which that seem very relevant are a person's gender as well as their education level. This paper uses a logistic regression to investigate the relattionship between these factors and a person's political preference

1 Introduction

You can and should cross-reference sections and sub-sections. We use R Core Team (2023) and Wickham et al. (2019).

The remainder of this paper is structured as follows. Section 2....

2 Data

2.1 Data Source

The data used in this paper is contained in the Cooperative Election Study 2022.(CITE) This study examines American's views on representatives, electoral experiences and elections. This data is available for free at the Harvard University Dataverse.

Surverys were used to collect data from a sample large enough that it may be assumed is representative of the entire American population. Importantly, this covered a wide range of constituencies and the sample was large enough such that the data for each constituency is enough to assume that it is representative of its whole.(CITE CES REPORT) There were

*Code and data are available at:https://github.com/Rahul-Uoft/election.git

60 research teams, and each purchased 1000 surverys to hand out to their respective constituencies. Sample matching was used to pick the 1000 people in which the surverys were to be administered. This is a process which is not truly random but has the benefit of being practical. In a truly randomly selected group, the contact information for some may not be readily available and therefore makes the data collectin process much more difficult. Sample matching involves selecting a sample at random, and anyone who does not fit the criteria to proceed(in this case having your contact information available) is then replaced by a person in the pool who is believed to have similar characteristics.

2.2 Data Analysis

The data is summarized in **?@fig-dataexplore**.

3 Model

The goal of our modelling strategy is twofold. Firstly,…

Here we briefly describe the Bayesian analysis model used to investigate… Background details and diagnostics are included in Appendix B.

3.1 Model set-up

Define y_i as the political preference of the individual and is 1 if the individual prefers Biden and 0 if the person prefers trump. Then gender_i is the individual's gender and education_i is the individual's education

$$
\begin{aligned}
y_i|\pi_i &\sim \text{Bern}(\pi_i) && (1)\\
\text{logit}(\pi_i) &= \alpha + \beta_1 \times \text{gender}_i + \beta_2 \times \text{education}_i && (2)\\
\alpha &\sim \text{Normal}(0, 2.5) && (3)\\
\beta_1 &\sim \text{Normal}(0, 2.5) && (4)\\
\beta_2 &\sim \text{Normal}(0, 2.5) && (5)\\
& && (6)
\end{aligned}
$$

We run the model in R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2022). We use the default priors from `rstanarm`.

3.1.1 Model justification

A logistic regression model was used in this analysis as the outcome variable is Binary. This is because we are considering whether someone prefers Biden or Trump. Whenever the outcome variable is Binary, a logistic regression model makes sense to help us understand it more. We used the default priors for our input variables.

A logistic regression model works by instead of considering an error value, which is done in linear regression, it considers a distribution for each of the inputs. The variability of these distributions inherently create the variability of the outcome which is associated to the error value we get in linear regression. The main advantage to this comes from the assumption made in linear regression models in which the error value is assumed to cancel out with each other(forming a normal distribution)

4 Results

Our results are summarized in Table 1.

5 Discussion

5.1 First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

5.2 Second discussion point

5.3 Third discussion point

5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Table 1: Explanatory models of flight time based on wing width and wing length

	First model
(Intercept)	−0.22
	(0.06)
genderMale	0.45
	(0.02)
educationHigh school graduate	0.07
	(0.06)
educationSome college	−0.31
	(0.06)
education2-year	−0.28
	(0.06)
education4-year	−0.61
	(0.06)
educationPost-grad	−0.94
	(0.06)
Num.Obs.	47 466
R2	0.037
Log.Lik.	−31 245.082
ELPD	−31 252.1
ELPD s.e.	55.6
LOOIC	62 504.2
LOOIC s.e.	111.1
WAIC	62 504.2
RMSE	0.48

Appendix

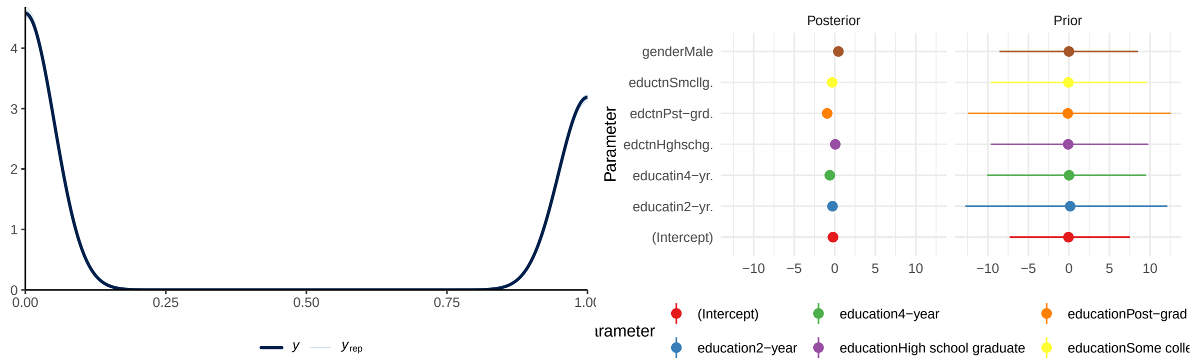
A Additional data details

B Model details

B.1 Posterior predictive check

In Figure 1a we implement a posterior predictive check. This shows...

In Figure 1b we compare the posterior with the prior. This shows...



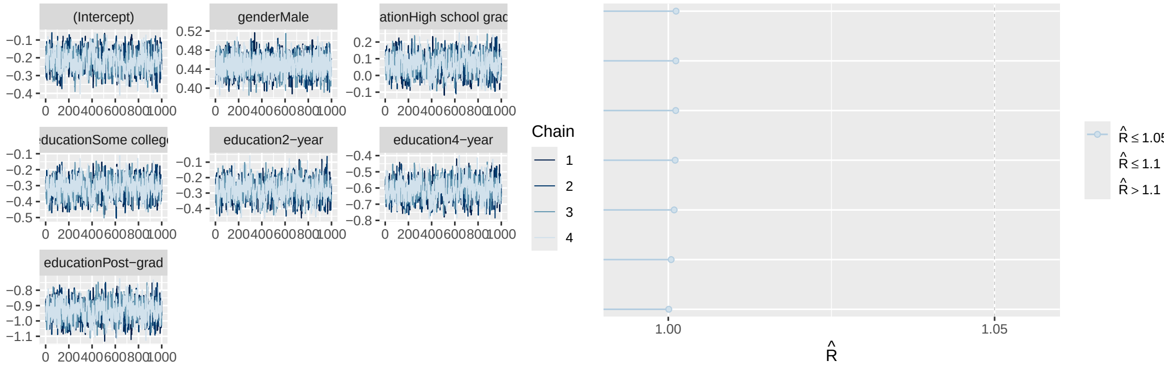
(a) Posterior prediction check (b) Comparing the posterior with the prior

Figure 1: Examining how the model fits, and is affected by, the data

B.2 Diagnostics

Figure 2a is a trace plot. It shows... This suggests...

Figure 2b is a Rhat plot. It shows... This suggests...



(a) Trace plot

(b) Rhat plot

Figure 2: Checking the convergence of the MCMC algorithm

References

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/.

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.