Investigating the correlation between a vote for Biden or Trump based on a persons education, gender, race and age.*

Fares Alkorani Owais Zahid

March 11, 2024

In this paper, we discuss the correlation between various statistics of the respondents to the CES2020 election survey, and whether they voted for Trump or Biden. The statistics are: age, education, gender and race.

1 Introduction

You can and should cross-reference sections and sub-sections. We use (**citeR?**) and (**rohan?**).

The remainder of this paper is structured as follows. Section 3....

2 Simulation

We first start with simulating the dataset, and recall that we are concerned with each voter's education level, gender, race and age. Our chosen sample size is 8000, and the categories for education and race have been named akin to the categories in the dataset. Thus, the header of the table obtained for the simulation is shown below:

`Table 1. Header for the political preferences dataset.`

*Code and data are available at: https://github.com/FFFiend/linear_model_investigation.

```
# A tibble: 6 x 5
  education     gender race                age supports_biden
  <chr>         <chr>  <chr>             <int> <chr>
1 Some college  Female Black                14 yes
2 College       Male   Two or more races    36 yes
3 < High school Female Black                77 yes
4 High school   Female White                80 yes
5 College       Female Middle Eastern       20 yes
6 High school   Female Asian                64 yes
```

3 Data

Some of our data is of penguins (**?@fig-bills**), from (**palmerpenguins?**).

Talk more about it.

And also planes (**?@fig-planes**). (You can change the height and width, but don't worry about doing that until you have finished every other aspect of the paper - Quarto will try to make it look nice and the defaults usually work well once you have enough text.)

Talk way more about it.

4 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.

4.1 Model set-up

Define y_i as the number of seconds that the plane remained aloft. Then β_i is the wing width and γ_i is the wing length, both measured in millimeters.

$$y_i|\mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma) \tag{1}$$
$$\mu_i = \alpha + \beta_i + \gamma_i \tag{2}$$
$$\alpha \sim \text{Normal}(0, 2.5) \tag{3}$$
$$\beta \sim \text{Normal}(0, 2.5) \tag{4}$$
$$\gamma \sim \text{Normal}(0, 2.5) \tag{5}$$
$$\sigma \sim \text{Exponential}(1) \tag{6}$$

We run the model in R (**citeR?**) using the `rstanarm` package of (**rstanarm?**). We use the default priors from `rstanarm`.

4.1.1 Model justification

We expect a positive relationship between the size of the wings and time spent aloft. In particular...

We can use maths by including latex between dollar signs, for instance θ.

5 Results

Our results are summarized in **?@tbl-modelresults**.

6 Discussion

6.1 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.

6.2 Second discussion point

6.3 Third discussion point

6.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

A Additional data details

B Model details

B.1 Posterior predictive check

In **?@fig-ppcheckandposteriorvsprior-1** we implement a posterior predictive check. This shows...

B.2 Diagnostics

C References