Forecasting the 2024 U.S. Presidential Election*

My subtitle if needed

Jason Yang Another author

November 2, 2024

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

1 Introduction

Overview paragraph

Estimand paragraph

Results paragraph

Why it matters paragraph

Telegraphing paragraph: The remainder of this paper is structured as follows. Section 2….

2 Data

2.1 Overview

We use the statistical programming language R (R Core Team 2023)…. Our data (Toronto Shelter & Support Services 2024)…. Following Alexander (2023), we consider…

Overview text

2.2 Measurement

Some paragraphs about how we go from a phenomena in the world to an entry in the dataset.

*Code and data are available at: [https://github.com/Jerryx2020/US_election_prediction/tree/main].

2.3 Outcome variables

Add graphs, tables and text. Use sub-sub-headings for each outcome variable or update the subheading to be singular.

Some of our data is of penguins (**?@fig-bills**), from Horst, Hill, and Gorman (2020).

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.

2.4 Predictor variables

Add graphs, tables and text.

Use sub-sub-headings for each outcome variable and feel free to combine a few into one if they go together naturally.

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 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 (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

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.

4 Results

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

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

Please don't use these as sub-heading labels - change them to be what your point actually is.

5.3 Third discussion point

5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

A Additional data details

B Model details

B.1 Idealized Methodology

In designing a forecast for the upcoming US presidential election with a $100,000 budget, our methodology focuses on creating a representative survey that captures the voting intentions of likely voters across the United States. The objective is to gather reliable, high-quality data that supports a well-informed forecast through effective sampling, data validation, and aggregation techniques.

Our target population is voters aged 18 and above, with a sample size of approximately 10,000 responses, 200 responses from each state, to ensure a statistically significance at the national level and within key demographic groups. To achieve balanced representation, we will employ stratified random sampling based on age, gender, race/ethnicity, education, region, and urban/rural status. This approach will involve weighting responses using recent census data and voter turnout statistics to address any demographic discrepancies.

For recruitment, we will combine online survey panels, social media outreach, and targeted telephone recruitment. These methods are chosen to increase accessibility for harder-to-reach demographics, such as older adults and rural residents. To encourage participation, we will offer modest incentives, such as a small monetary compensation or entry into a prize draw. All data collection will be conducted through a reliable survey platform—Google Forms.

The survey itself will be concise and straightforward, focusing on voting intentions, key issues, and demographic information. For data validation, in the survey, we will include attention-check questions as respondents are very likely to complete the form for an entry into a prize draw. We will also conduct multiple waves of the survey, we will adopt a "poll-of-polls" approach, aggregating responses from different waves while applying weights based on relevance.

To mitigate biases such as non-response bias, social desirability bias, and order bias, our team will implement several strategies. We will send reminders to non-respondents to encourage participation, emphasizing the prize draw incentives to increase response rates. Survey questions will be carefully crafted to maintain neutral wording and avoid sensitive topics that might discourage honest responses. Additionally, within each wave of polls, we will randomize the order of questions to minimize the impact of question order on responses. These measures aim to enhance the reliability and accuracy of our survey results.

Ethically, our approach prioritizes respondent privacy and data security. We will ensure that participants' responses remain anonymous and that all data is handled securely. Transparency is also essential; respondents will be informed about the survey's sponsorship, methodology, and data use

The survey will be implemented on Google Forms, with a link provided in the appendix. We also included the questions in the appendix for better understandings.

B.2 Idealized Survey

The proposed survey is designed using Google Forms,https://docs.google.com/forms/d/e/1FAIpQLSeHHeKaryN
ZgXwQYkNw9jBdxIbgeXWVj_t1ob2w/viewform?usp=sf_link

B.3 Survey Copy

1. **Who do you intend to vote for in the 2024 U.S. Presidential Election?**

- Kamala Harris (Democratic Party)
- Donald Trump (Republican Party)
- Undecided
- Other:

2. **How certain are you about your choice?**

- Very Certain
- Somewhat certain
- Unsure

3. **Which of the following issues are most important in deciding your vote? (Select up to 3)**

- Economy and jobs
- Healthcare
- Climate Change
- Education
- Immigration
- National Security
- Social justice and Equality
- Abortion rights
- Gun control
- Other:

4. **What is your age?**

- 18-24
- 24-34
- 35-44
- 45-54
- 55-64
- 65 or older

5. **What is your gender?**

- Male
- Female
- Non-binary/Other
- Prefer not to say

6. **What is your race/ethnicity? (Select all that apply)**

- White
- Black or African American
- Hispanic or Latino
- Asian
- Native American or Alaska Native
- Native Hawaiian or Other Pacific Islander
- Other:

7. **In which U.S. state do you currently reside**

- _____

8. **If you voted in 2020, who did you vote for?**

- Joe Biden
- Donald Trump
- Other:

References

Alexander, Rohan. 2023. *Telling Stories with Data*. Chapman; Hall/CRC. https://tellingstorieswithdata.com/.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "rstanarm: Bayesian applied regression modeling via Stan." https://mc-stan.org/rstanarm/.

Horst, Allison Marie, Alison Presmanes Hill, and Kristen B Gorman. 2020. *palmerpenguins: Palmer Archipelago (Antarctica) penguin data*. https://doi.org/10.5281/zenodo.3960218.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Toronto Shelter & Support Services. 2024. *Deaths of Shelter Residents*. https://open.toronto.ca/dataset/deaths-of-shelter-residents/.