A Prediction of the 2024 U.S. Election*

Prediction Based on MRP Analysis of Harris Lost

Shanjie Jiao Aman Rana Kevin Shen

October 28, 2024

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

1 Introduction

1.1 Overview paragraph

The United States, as the world's largest economy and most influential country, will hold its presidential election on November 5, 2024. Despite facing domestic challenges such as inflation, low employment rates after the pandemic, and increasing political polarization due to conflicts between the Democratic and Republican parties, which have accelerated internal tensions.As of October 2024, former President Donald Trump has secured the Republican Party nomination. However, after assassination happened on July 14, 2024, many of his supporters have become even more committed to his legacy. Meanwhile, President Joe Biden withdrew from the election on July 21, leading to the nomination of Kamala Harris as the Democratic candidate. These events have made the 2024 U.S. election increasingly complex and unpredictable, further intensifying uncertainty about the future.

1.2 Estimand paragraph

In this paper, multiple linear regression will be used to fit the dataset provide by Five Thirty Eight[use citation] to estimate Trump's election rate.

*Code and data are available at: [https://github.com/Jie-jiao05/2024_US_Election.git).

1.3 Results paragraph

1.4 Why it matters paragraph

Due to the United States' hegemonic position and unparalleled global influence, the outcome of the U.S. election will serve as a guiding light for global developments over the next four years, significantly impacting international security, trade, cooperation, and geopolitics. This article aims to provide a clearer analytical framework for political scientists, the general public, journalists, corporate strategists, and anyone globally concerned with the U.S. election.

Telegraphing paragraph: The remainder of this paper is structured as follows. Section 2 we will discuss the MLR used in the prediction and the result of the MLR with its predictor variables and the response after transformation. **?@sec-dis** the shortcomings of the study and areas for improvement will be described. **?@sec-A** A Deep Dive to a pollster will be conducted, and **?@sec-B** We will given a idealized survey and methodology

2 Data

2.1 Overview

We use the statistical programming language R (R Core Team 2023) to analysis the data which are extracted from the 2024 U.S. General President Election collected and published by FiveThirtyEight (**citation?** later). This paper will conduct a Multiple Linear Regression to do further prediction. In this model, we take numeric_grade, pollscore, sample_size as predictors, and pct as response.

2.2 Data Cleaning

To Simplify the dataset we extract the data of the pollater_id, pollster, methodology, numeric_grade, pollscore, sample_size, pct, and transform Party and Answer into categorical (0=DEM, 1=REP and 0 = Harris, 1=Trump). In oreder to make the pollster more reliable we will only analysis that transparency are over 3 to increase the liability of this survey.

Overview text

2.3 Measurement

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

2.4 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 (Figure 1), from Horst, Hill, and Gorman (2020).

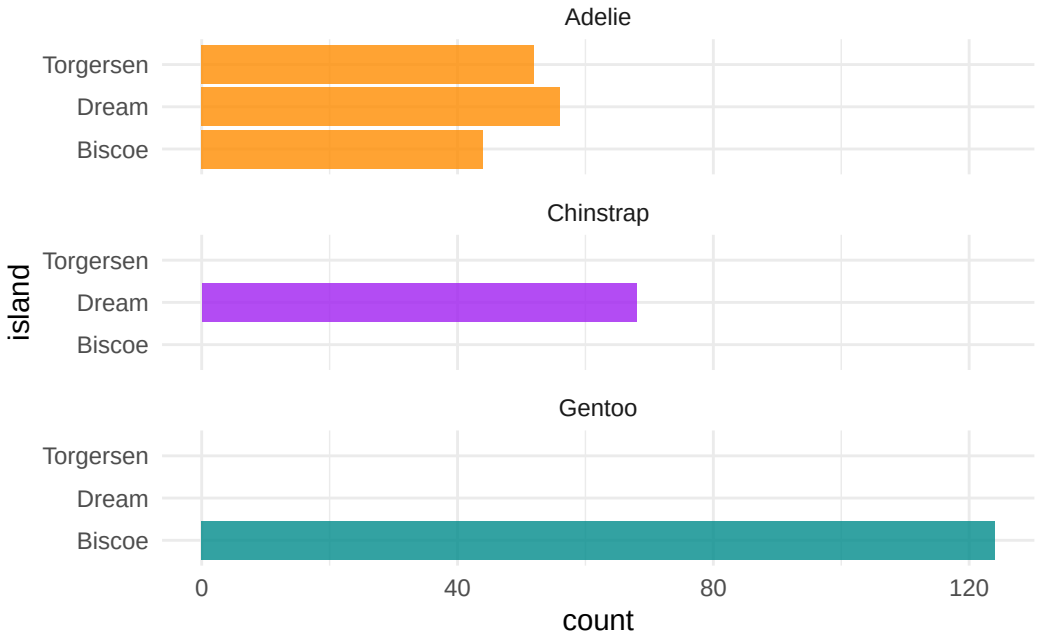


Figure 1: Bills of penguins

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.

2.5 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 Table 1.

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

	First model
(Intercept)	1.12
	(1.70)
length	0.01
	(0.01)
width	−0.01
	(0.02)
Num.Obs.	19
R2	0.320
R2 Adj.	0.019
Log.Lik.	−18.128
ELPD	−21.6
ELPD s.e.	2.1
LOOIC	43.2
LOOIC s.e.	4.3
WAIC	42.7
RMSE	0.60

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.

A Appendix

A.1 B: Idealized Survey Methodology – $100K Budget

A.1.1 Introduction

In this appendix, we outline an idealized survey methodology for forecasting U.S. presidential elections within a budget of up to $100,000. The proposed design aims to maximize accuracy and cost-efficiency by strategically selecting sample populations, targeting key demographic groups, and effectively aggregating results.

A.1.2 Sampling Strategy

The ideal survey methodology employs a stratified random sampling approach, given the diversity of the American population as the target group. This method is designed to better analyze the varying effects of different races, age groups, genders, and education levels on voting trends by distributing the sample into multiple subgroups. By categorizing these key variables, this approach ensures more concise and clearer statistical results, which in turn facilitates in-depth analysis. The representative sample is listed below:

- **Sampling Frame**: For the information security, valid entitlement holders registered on the voter list and registered with the Census Bureau will receive the survey.
- **Sampling Method**: Stratified random sampling will be implemented, ensuring better analyze the differential impact of different key factors, including races, ages, genders, and levels of education on voting trends.
- **Sample Size**: The poll is expected to include responses from 10,000 individuals. Assuming nonresponse bias is 10%, we estimate the final adjusted sample size of approximately 9,000.
- **Geographical Distribution**: The survey will be conducted across all 50 states and the District of Columbia. Considering that traditional "red" and "blue" states are less likely to change political alignments, an increased focus will be placed on polling within swing states.

A.1.3 Recruitment Plan

This survey will conduct recruiting in following methods:

- **Online recruitment**: Advertisements and official information about the survey will be disseminated through various search engines (including Google and Firefox) and social media platforms such as X and Instagram. All advertisements will contain direct links to the Google Forms survey.

- **Phone recruitment**: Valid phone numbers will be contacted by staff to conduct the survey using the same set of questions as the online survey. For privacy and security, no conversations will be recorded, and the calling system will automatically dial numbers without disclosing them to staff members.
- **In person recruitment**: A tent will be set up in key high-traffic regions in swing states to conduct survey in-person.
- **Gift incentive**: To increase participation, all respondents will receive a specially designed pin, estimated to cost around $2 each. This incentive is intended to boost engagement and encourage broader participation in the survey.

A.1.4 Budgect Allocation

- **Online Recruitment Cost**: $20,000 allocated for advertising across various platforms (search engines and social media).
- **Phone Recruitment Cost**: $20,000 allocated for hiring staff to conduct phone surveys and cover the associated data usage charges.
- **In-Person Recruitment Cost**: $22,000 allocated for hiring field staff and setting up in-person recruitment locations.
- **Gift Cost**: $20,000 allocated for the total value of participant gifts
- **Data Analysis and Quality Control Cost**: $6,000 allocated for employing technical staff, purchasing software licenses, and utilizing required tools.
- **Administrative Cost**: $2,000 allocated for overall management, including miscellaneous expenses and logistical support.

Total Budget: $100,000

A.1.5 Survey Design

This survey will only collect data that are accurate and unbigoted. Meanwhile, personal information will not be contained.

- **Close-ended Question**: Usage of multiple choice can make sure the answer is straightforward and easier for analysis.
- **Response Option**: Most questions will only contain option from A, B and "prefer not to say", all the option will be a neutral response.
- **Insensitive Information**: This survey will only contain personal information only and merely age, gender, race, education background, income, living state.

A.1.6 Data Process

To ensure data we collected are valid and significant, we will conduct following methods to test our data.

- **One-times Submission**: For online surveys, only one submission will be allowed per unique IP address to prevent multiple entries that could skew results. The same principle will apply to phone recruitment, ensuring each number is only surveyed once.
- **Missing Data**: During the data cleaning process, responses with missing values (N/A entries) will be excluded to maintain the integrity and reliability of the analysis.
- **Swing States Data Process**: Data collected from swing states will be assigned a slightly higher weight in the estimation process, given their critical influence on election outcomes.
- **Stratified Analysis**: To gain insights into voter preferences within different demographic subgroups, stratified analyses will be conducted, focusing both on overall trends and trends within specific subgroups.
- **Poll Aggregation Approach**: The aggregated data will be weighted based on sample size (with larger samples receiving greater weight), and past reputable polls will be incorporated, giving priority to highly rated pollsters to strengthen the overall analysis.

A.1.7 Conclusion

The processes outlined in this survey aim to provide valid and meaningful insights into the upcoming U.S. 2024 Presidential Election. Working within a limited budget, we have carefully designed the survey and data processing methods to achieve reliable and comprehensive analyses, both at the subgroup level and in the overall population.

Sample survey could find here: Surey Link.

B Model details

B.1 Posterior predictive check

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

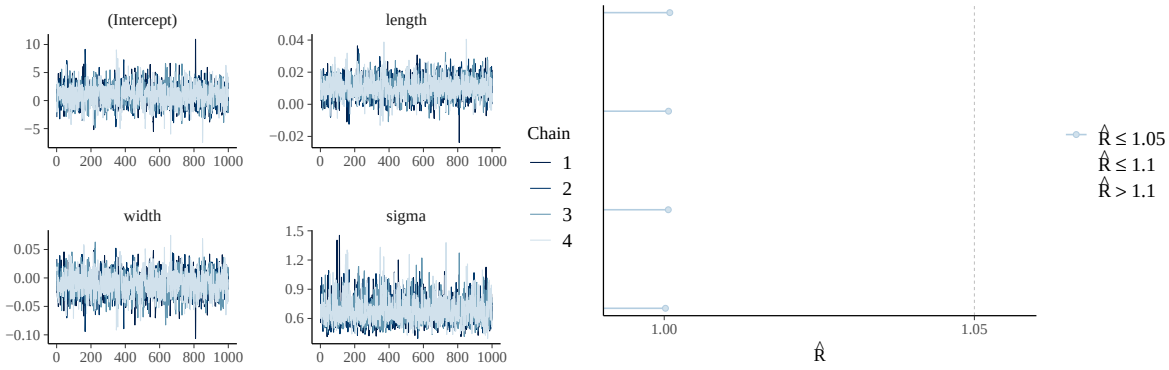
In **?@fig-ppcheckandposteriorvsprior-2** we compare the posterior with the prior. This shows…

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/.

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/.