Forecasting the 2024 US Presidential Election*
My subtitle if needed

Tina Kim   David Flores   Kevin Shao

October 16, 2024

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

1 Introduction

Overview paragraph: Provide a brief overview of the context of the upcoming US presidential election and the importance of forecasting its results.

Estimand paragraph: Define the estimand clearly (e.g., predicting the probability of a candidate winning the election based on poll data).

Results paragraph: Summarize the key findings of the model, highlighting its predictive accuracy and implications.

Why it matters paragraph: Explain the broader significance of accurately forecasting election results for politics, society, and policymaking.

Telegraphing paragraph: The remainder of this paper is structured as follows: Section 2 details the data and measurement process; Section 3 covers model development and results; Section 4 discusses implications and future steps. The remainder of this paper is structured as follows. Section 2....

2 Data

2.1 Overview

We obtained the "Presidential general election polls" dataset from FiveThirtyEight (FiveThirtyEight 2024) and used the statistical programming language R (R Core Team 2023) to analyze this data.

---

*Code and data are available at: https://github.com/DavidFJ207/USPresidentialForecast

Person A: Provide an overview of the dataset's context, sources, and key variables.

Our objective was to clean, organize, and analyze polling data for the U.S. presidential general election. Below is an outline of the steps we took to achieve this goal:

2.1.1 1. Cleaning Raw Poll Data

We began by identifying the three most frequent pollsters within the dataset, focusing on these pollsters to improve data reliability and relevance. Missing values were assessed for each pollster, and the pollster with the fewest missing values was selected to ensure more complete data for further analysis. This refined dataset was then saved for the next stages of the process.

2.1.2 2. Organizing Poll Links by State

Next, we organized the polling data by state, ranking the links for each state to facilitate easier access. To ensure clarity, missing values were shifted to the end of each state's record, and unnecessary empty columns were removed. This resulted in a more structured and user-friendly dataset, which was stored for future use.

2.1.3 3. Extracting and Merging Poll Data from Links

Following that, we extracted and merged additional polling data from Google Sheets links included in the dataset. After confirming the validity of these links, we cleaned the data by removing unwanted entries and filled in any missing values where possible. This updated dataset was then merged with the original dataset to create a more comprehensive compilation of polling data.

2.1.4 4. Final Organization and Refinement

In the final stage, we further refined the dataset by removing unnecessary rows (such as totals and empty states) and renaming specific entries for clarity (e.g., changing "State" to "National"). Columns with excessive missing values were also eliminated to ensure that only the most reliable and complete data remained. This final, polished dataset was saved, ready for subsequent analysis and reporting.

2.2 Measurement

Describe how the real-world phenomenon (voter preference) is translated into a dataset entry.

Person C: Discuss the survey design, the process of data collection, and potential biases or errors in measurement.

2.3 Outcome variables

Add graphs, tables, and text for each outcome variable. Use subsections as needed.

Some of our data is (**?@fig-outcome**), from (**source?**)

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 to describe key predictors.

Person B:: Visualize variable distributions (e.g., histograms, boxplots), and justify the inclusion of predictors in the model.

3 Pollster Methodology Overview (Appendix A)

Append a detailed review of a chosen pollster's methodology, survey techniques, strengths, and weaknesses.

4 Model

4.1 Model Development

The goal of our model is to forecast the popular vote outcome of the 2024 US presidential election.

Person B: Choose a linear or generalized linear model. Justify the selection based on your research goal and data structure.

4.2 Model Set-up

Define the model mathematically and contextually (e.g., linear model predicting vote share).

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

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`.

Person B: Define all model components and assumptions.

4.3 Model Justification

Our results are summarized in Table **??**. Discuss the implications of the findings and their predictive accuracy.

5 Results

Our results are summarized in Table **??**.

Person B: Visualize the model's results and include any performance metrics (e.g., RMSE, test/train split).

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

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

6.3 Third discussion point

Discuss what the model reveals about the election forecast and its potential impact on understanding voting behavior.

Person C: Discuss limitations of the model and areas for further improvement.

6.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

.1 Appendix A: Pollster Methodology Overview

Person A: Provide a detailed methodology review for the chosen pollster, including survey design, population sampling, non-response handling, etc.

.2 Appendix B: Idealized Survey Design for $100K Budget

Person C: Develop and describe an ideal survey design for forecasting the election with a $100K budget, and include survey link.

.3 Additional Data & Model Details

Include any technical details on data cleaning, model diagnostics, and posterior checks.

A Additional data details

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

?@fig-stanareyouokay-1 is a trace plot. It shows... This suggests...

?@fig-stanareyouokay-2 is a Rhat plot. It shows... This suggests...

Checking the convergence of the MCMC algorithm

References

FiveThirtyEight. 2024. "Dataset: US Presidential General Election Polls." https://projects.fivethirtyeight.com/polls/data/president_polls.csv.

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/.