Forecasting the 2024 U.S. Presidential Election: A Poll-Based Linear Model Approach

Aamishi Avarsekar Divya Gupta Gauravpreet Thind

October 22, 2024

This research aims to develop a generalized linear model to predict the outcome of the 2024 U.S. Presidential Election. The model aggregates polling data from various sources via the poll-of-polls method to provide a forecast of voter behavior regarding Kamala Harris, the Democratic Candidate, and Donald Trump, the Republican Candidate. This paper presents our exploratory data analysis (EDA) to clean and prepare the dataset followed by the predictive model and its results, which project a Kamala Harris (?) win by X%. The paper further presents an idolized survey to test these findings and discusses the methodology to run it using a USD $100K budget.

1 Introduction

The 2024 U.S. Presidential Election is set to take place on 5th November this year. This year, it is dominated by the competition between the Democratic candidate, Vice President Kamala Harris and her Vice President pick, Tim Walz, and the Republican candidate, former President Donald Trump, and his Vice President pick, JD Vance. Using regularly updated polling data allows one to forecast the outcome of the elections ahead of time and draw meaningful conclusions about voter behavior and the US electoral system. Forecasting the election also allows presidential candidates to recognise gaps in their campaigns and compels them to re-assess their platform to one that is better suited and favorable for the majority voting population.

Accurately forecasting the election outcomes is a dynamic and challenging process which requires one to assess voter behavior (which may vary county by county based on a plethora of factors such as age, sex, income levels and race), polling methodologies, etc. In this paper, we aim to develop a general linear predictive model that forecasts the outcome of the 2024 elections through a polls-of-polls approach. This approach aims to aggregate the results of multiple polls to develop a more reliable estimate of voter behavior across different sample sizes, counties, and demographic features. To develop this model, we use the publicly available polling data retrieved from FiveThirtyEight *(insert info here)*. Our analysis focused on data collected between July and October 2024, and includes key voter demographic variables such as state of residence, pollster rating, and sample size.

This paper is structured such that Section 2 describes the dataset, including high-level data cleaning through our exploratory data analysis, variables used in the analysis, summary of statistics and necessary data visualizations. Section 3 presents the predictive model, its results, strengths and limitations. This allows the researchers to discuss next steps to improve the model and conclude our findings in Section 4. Appendix A presents an analysis of Kamala Harris' campaign methodology in this election, and Appendix B presents a sample survey we could run to assess the efficacy of our model along with a potential allocation of a USD $100K budget to run this survey.

2 Data

2.1 EDA

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

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 .

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.

Appendix

A Additional data details

B Model details

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/.

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.