Predicting the 2024 US Presidential Election: A Polling-Based Forecast*

My subtitle if needed

Tianrui Fu Yiyue Deng

October 25, 2024

This paper forecasts the 2024 US Presidential election outcome using polling data from [insert pollster name]. By applying simple and multiple linear regression models, we analyze the effect of polling factors, including sample size, poll score, and transparency, on support percentages for key candidates. Our findings suggest significant relationships between these variables, providing an evidence-based approach to predicting election outcomes. We further explore methodological strengths and weaknesses and propose an ideal polling survey methodology. This work highlights the potential for data-driven insights into political forecasting.

Table of contents

*Code and data are available at: https://github.com/RohanAlexander/starter_folder.

1 Introduction

Overview paragraph

Estimand paragraph

Results paragraph

Why it matters paragraph

Telegraphing paragraph: The remainder of this paper is structured as follows. Section 5….

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.

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[].

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 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

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

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

	Model Summary	
	Model by glm	Model by bayes
(Intercept)	−524.927*** [−773.481, −276.373]	−590.212 [−814.191, −354.733]
numeric_grade	−14.386 [−43.487, 14.716]	−3.356 [−11.443, 5.059]
transparency_score	0.903 [−0.348, 2.153]	−0.164 [−0.881, 0.569]
pollscore	−3.028 [−32.532, 26.476]	1.121 [−5.151, 7.339]
Beacon/Shaw	−2.833 [−12.913, 7.246]	0.976 [−1.509, 3.389]
Christopher Newport U.	−9.186* [−18.242, −0.130]	−5.294 [−10.778, 0.269]
CNN/SSRS	−4.208 [−10.838, 2.423]	−1.343 [−4.986, 2.341]
Data Orbital	−3.602 [−12.297, 5.093]	−0.426 [−6.114, 5.093]
Echelon Insights	−3.313 [−7.738, 1.113]	0.389 [−2.466, 3.257]
Emerson	1.010 [−6.052, 8.072]	1.561 [−0.559, 3.788]
Ipsos	−7.483** [−13.082, −1.883]	−3.927 [−6.355, −1.584]
Marist	0.162 [−4.587, 4.911]	1.111 [−0.987, 3.216]
Marquette Law School	−3.213 [−12.696, 6.270]	−0.561 [−3.603, 2.499]
MassINC Polling Group	−7.280*** [−10.601, −3.958]	−6.566 [−8.857, −4.172]
McCourtney Institute/YouGov	−1.354 [−10.616, 7.908]	−1.745 [−7.385, 3.817]
Muhlenberg	−2.040 [−10.915, 6.834]	0.196 [−5.405, 5.867]
Quinnipiac	−1.513 [−10.660, 7.633]	−0.189 [−3.441, 3.117]
Selzer	−2.952 [−15.951, 10.046]	−0.220 [−5.949, 5.492]
Siena	−9.684*** [−13.295, −6.072]	−7.344 [−10.509, −4.019]
Siena/NYT	−2.174 [−20.461, 16.113]	0.976 [−3.002, 5.133]
Suffolk	−4.753 [−10.985, 1.479]	−3.290 [−6.325, −0.369]
SurveyUSA	−5.810 [−20.672, 9.053]	−2.163 [−5.651, 1.268]
SurveyUSA/High Point University	−5.876 [−23.153, 11.401]	−0.224 [−5.003, 4.990]
The Washington Post	0.576 [−10.255, 11.407]	2.315 [−1.070, 5.912]
University of Massachusetts Lowell/YouGov	−6.065 [−13.331, 1.201]	−3.502 [−7.054, 0.165]
Washington Post/George Mason University	−8.943*** [−14.000, −3.886]	−4.783 [−8.412, −1.163]
YouGov	−2.338 [−10.874, 6.199]	−0.764 [−3.385, 1.935]
YouGov Blue	−1.128 [−11.743, 9.488]	0.335 [−5.326, 5.727]
End Date	0.030*** [0.019, 0.042]	0.032 [0.021, 0.044]
U. North Florida		1.990 [−3.834, 7.845]
YouGov/Center for Working Class Politics		−2.360 [−8.276, 3.399]
Num.Obs.	492	492
R2	0.332	0.318
R2 Adj.		0.258
AIC	2575.9	
BIC	2701.8	
Log.Lik.	−1257.932	−1264.371
ELPD		−1289.0
ELPD s.e.		25.3
LOOIC		2577.9
LOOIC s.e.		50.6
WAIC		2575.6
RMSE	3.12	3.38

+ p \num{< 0.1}, * p \num{< 0.05}, ** p \num{< 0.01}, *** p \num{< 0.001}

[a] This table shows the regression models with custom variable names.

5.3 Third discussion point

5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

.1 Appendix B: Idealized Survey Methodology

.1.1 Introduction

The budget for this survey is $100,000, aimed at predicting the outcome of the 2024 United States Presidential Election. This methodology includes stratified sampling, multi-platform recruitment, data validation, and multi-wave data aggregation to ensure representative and reliable data.

.1.2 Sampling Approach: Stratified Random Sampling

To obtain a representative sample, we use stratified random sampling with a sample size of 5,000 respondents. Sampling is stratified by age, gender, education, and geographic region to ensure broad coverage across voter demographics. The age groups include 18-29, 30-44, 45-64, and 65 and above. Gender is categorized as male, female, and other, while education is classified as high school or below, bachelor's degree, and master's degree or above. The geographic region includes all U.S. states. Stratified sampling ensures that the sample accurately reflects voter characteristics, providing a solid foundation for subsequent analysis.

.1.3 Recruitment: Multi-Platform and Interactive Engagement

Recruitment is conducted through a multi-platform strategy to ensure wide coverage among voters. Targeted ads are deployed on Google, Facebook, and Twitter to attract respondents with specific demographic characteristics. The ad content is concise and highlights anonymity and the research purpose to encourage participation. In addition, we use Random Digit Dialing (RDD) to contact older adults and residents in remote areas who may be less accessible online, ensuring their participation in the survey. To increase the completion rate, a small reward system is implemented, with one out of every 100 participants receiving a $5 to $10 incentive.

.1.4 Survey Platform: Google Forms and Phone Outreach

Google Forms is used as the primary data collection platform, enabling respondents to complete the survey on a computer or mobile device with ease. We have set Google Forms to restrict submissions to one per Google account to prevent duplicate responses. For those who may not have easy online access, telephone outreach will be conducted, especially targeting older adults and rural voters, to ensure the diversity and inclusivity of the sample.

.1.5 Data Validation: Post-Stratification Weighting

To ensure data quality, post-stratification weighting will be applied during the analysis phase based on demographic characteristics. This adjustment aligns the sample structure with the national voter distribution, reducing bias and increasing the accuracy of our findings. Multi-layered data validation measures enhance the reliability and representativeness of the results.

.1.6 Poll Aggregation: Multi-Wave Polling and Adjustments

To track changes in voter sentiment over time, we will conduct multiple waves of data collection and aggregation. The survey will be administered in multiple rounds throughout the election cycle, with each round spaced 3-4 weeks apart. Each wave of data will be collected and analyzed independently, then aggregated using weighted averages to smooth out single-instance fluctuations and help identify long-term trends, providing a reliable basis for predictions.

.1.7 Budget Allocation: Phased Spending and Testing

To maximize budget efficiency, we will use a phased spending strategy. An initial 30% of the advertising budget will be allocated to testing across platforms to determine effectiveness, with the remaining 70% directed to the most successful channels. The budget breakdown is as follows: social media advertising ($40,000) for broad and targeted outreach, phone outreach ($20,000) for RDD calls to less accessible populations, small rewards ($15,000) to encourage completion, and data cleaning and analysis ($25,000) for validation, weighting, and aggregation across multiple waves.

.1.8 Survey Content and Link

Survey Title: 2024 United States Presidential Election Survey

Survey Link: https://forms.gle/EzyHp3zuX8Cu6Ep8A

Survey Questions:

1. What is your age?

o 18-29

o 30-44

o 45-64

o 65 and above

2. What is your gender?

o Male

o Female

o Other

o Prefer not to say

3. What is your highest level of education?

o High school or below

o Bachelor's degree

o Master's degree or above

4. Which state do you currently reside in?

(Dropdown menu with state names)

5. Do you plan to vote in the 2024 election?

o Yes

o No

o Not sure

6. If the election were held today, which candidate would you be more likely to support?

o Candidate A (Democratic Party)

o Candidate B (Republican Party)

o Other

o Undecided

7. Which of the following issues would have the greatest impact on your decision in the 2024 election? (Select one)

o Economic stability and growth

o Access to quality healthcare

o Education reform and funding

o Environmental sustainability and climate action

o Addressing social and racial inequalities

o Other (please specify): _____

8. If a future candidate proposed a policy that aligns perfectly with your concerns, would you consider changing your voting intention?

o Yes

o No

o Not sure

9. If you would like to participate in the reward draw, please provide your email address below. Your contact information will only be used to notify winners. We will reach out to winners via email.

Email: _____

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 algo-
rithm

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/.

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/.