Prediction about 2024 U.S. Presidential Election Outcome Using Linear Modeling*

My subtitle if needed

Yunkai Gu Anqi Xu Yitong Wang

October 25, 2024

This study aims to forecast the outcome of the 2024 U.S. Presidential Election using aggregated polling data and a linear modeling approach. The study also includes an analysis of a selected pollster's methodology and an idealized survey design for forecasting elections within a limited budget.

Table of contents

*Code and data are available at: https://github.com/Kylie309/2024-US-election-prediction.

1 Introduction

The U.S. presidential election is a critical event that attracts substantial public and media interest. Various tools and methodologies have then been used for making predictions about the final electoral outcome.
Curious about the election outcome as well, this paper employs the "poll-of-polls" approach to make predictions, which aggregates multiple polls to minimize biases and enhance prediction accuracy. A linear model is built to forecast the winner of the 2024 U.S. Presidential Election. In addition, in-depth analysis on certain pollster's methodology is also included, discussing their sampling approach, strengths, and weaknesses.

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.

2.3 Outcome variables

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

Plot basic relationship

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 percentage vote for Trump. Then x_{i1} is the sample size and x_{i2} is the pollster.

$$y_i = \alpha + \beta_i x_{i1} + \gamma_i x_{i2}$$

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`.

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.

$$
\begin{aligned}
y_i|\mu_i, \sigma &\sim \text{Normal}(\mu_i, \sigma) && (1)\\
\mu_i &= \alpha + \beta_i + \gamma_i && (2)\\
\alpha &\sim \text{Normal}(0, 2.5) && (3)\\
\beta &\sim \text{Normal}(0, 2.5) && (4)\\
\gamma &\sim \text{Normal}(0, 2.5) && (5)\\
\sigma &\sim \text{Exponential}(1) && (6)
\end{aligned}
$$

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.

Posterior predictive checks for national spline model:

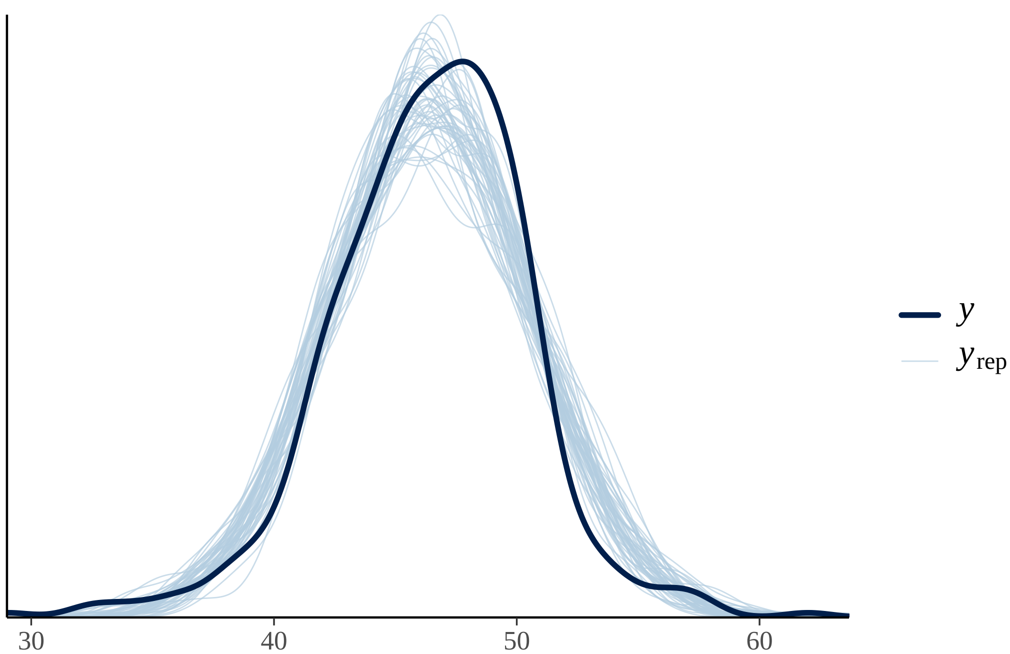
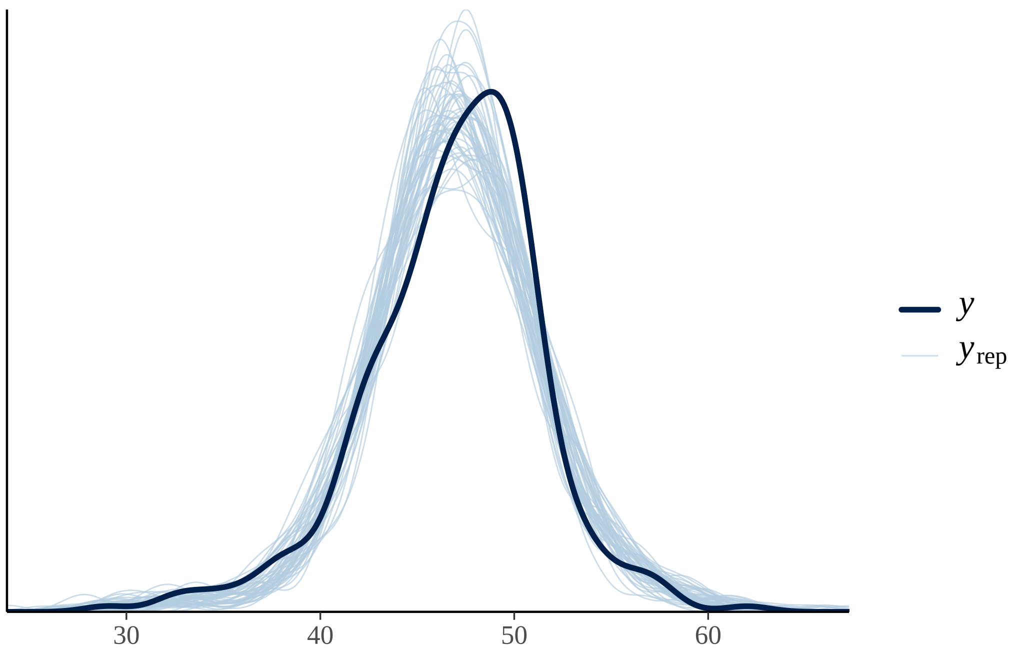


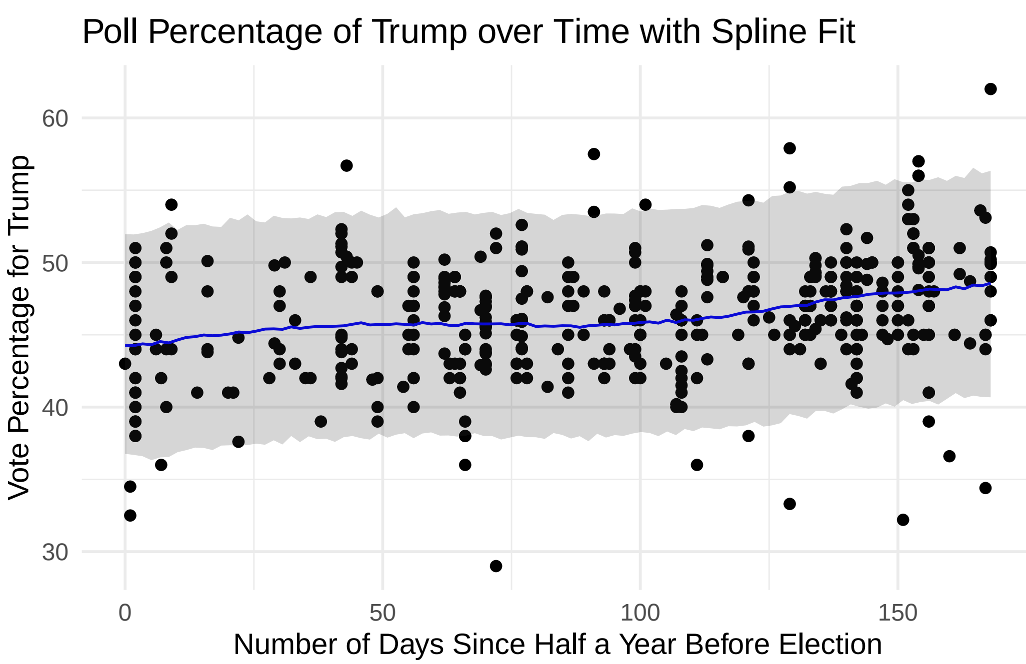
Table 1: Explanatory models of Trump Percentage Vote based on Sample Size and Pollster

	Model 1	Model 2
(Intercept)	44.27	45.59
ns(end_date_num, df = 5)1	1.19	0.21
ns(end_date_num, df = 5)2	1.60	2.92
ns(end_date_num, df = 5)3	3.10	1.85
ns(end_date_num, df = 5)4	5.58	6.28
ns(end_date_num, df = 5)5	3.17	2.94
stateCalifornia		−9.85
stateFlorida		3.62
stateGeorgia		0.24
stateIndiana		7.34
stateMaryland		−12.69
stateMassachusetts		−15.02
stateMichigan		−1.79
stateMinnesota		−1.91
stateMissouri		6.07
stateMontana		7.55
stateNebraska CD-2		−6.10
stateNevada		−0.36
stateNew Hampshire		−5.26
stateNew Mexico		−3.40
stateNew York		−5.42
stateNorth Carolina		−0.25
stateOhio		1.75
statePennsylvania		−1.42
stateSouth Dakota		9.12
stateTexas		1.24
stateVirginia		−4.38
stateWisconsin		−1.45
Num.Obs.	472	307
R2	0.089	0.603
R2 Adj.	0.068	0.557
Log.Lik.	−1310.219	−740.269
ELPD	−1316.8	−768.5
ELPD s.e.	21.8	11.9
LOOIC	2633.7	1537.0
LOOIC s.e.	43.6	23.9
WAIC	2633.7	1533.1
RMSE	3.87	2.67

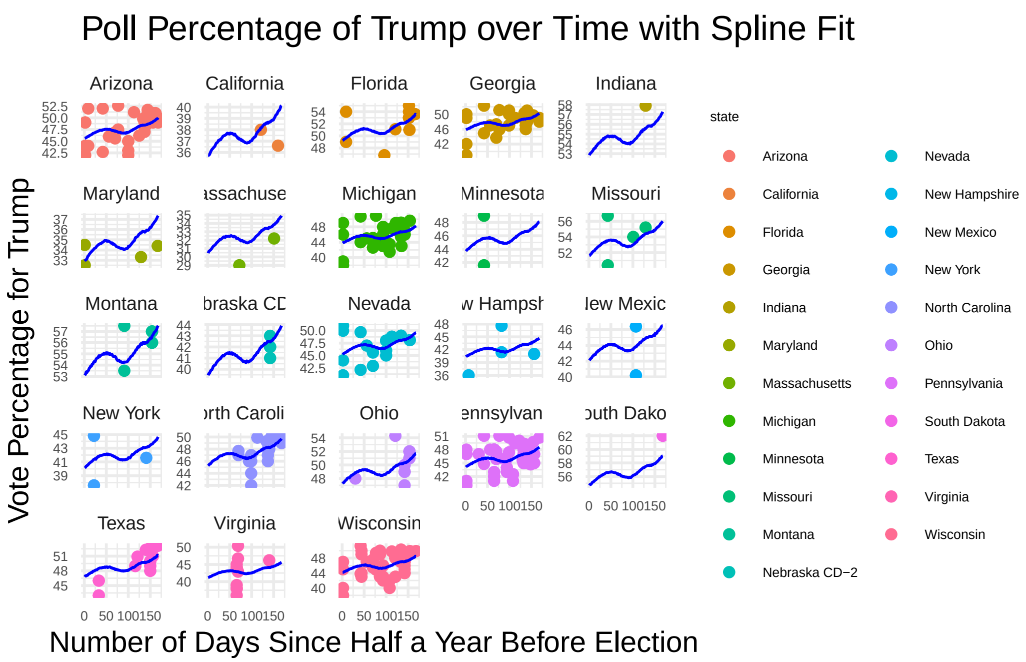
Posterior predictive checks for state spline model:



Predict posterior draws and spline fit for votes nationally



Predict posterior draws and spline fit for votes in each state



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 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

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/.