Red, Blue, or Purple? 2024's Battleground States Tell a Story*

Pennsylvania, Arizona, and Nevada Become Game Changers as Trump Gains Ground

Yingke He Ziheng Zhong

October 31, 2024

This paper

Table of contents

*Code and data are available at: https://github.com/iJustinn/Election_Prediction.

1 Introduction

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 Data 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.

2.4 Predictor variables

Some explanations

2.4.1 State

State (state)

2.4.2 Pollster

Pollster (pollster)

2.4.3 Candidate Name

Candidate Name (candidate_name)

2.4.4 Percentage of Votes

Percentage of Votes (pct)

2.4.5 Sample Size

Sample Size (sample_size)

2.4.6 Days to Election

The variable days to election represents the days remaining from the poll's end date to Election Day. It is calculated using the end_date variable, by assuming the election date is November 5 2024.

3 Model

To predict the actual election vote share for each candidate in each state while accounting for variations by pollster and other poll-specific factors. Background details and diagnostics are included in Appendix B.

3.1 Model set-up

A Hierarchical Bayesian Model is utilized to predict the actual election vote for each candidate in each state while accounting for variables by pollster.

The model prediction utilizes the following predictor variables:

- **State** (`state`): Include as a categorical term to capture regional variations.
- **Pollster** (`pollster`): Include as a categorical variable for different polling effects.
- **Candidate Name** (`candidate_name`):Include as a categorical feature or model separately for each candidate.
- **Percentage of Votes by Poll in State** (`pct`): Use as a primary predictor of actual vote share, with a smooth term to allow for non-linear effects.
- **Sample Size** (`sample_size`): Include as a predictor or weight to reflect poll reliability.
- **Days to Election** (`end_date`):capture trends in support leading up to the election.

Let:

y_{ijk} : The target variable, representing the actual election vote share for candidate k in state

pct_{ijk} : The observed polling percentage for candidate k in state i by pollster j.

sample_size_{ijk} : The sample size of the poll, which helps in weighing the poll reliability.

$\text{days\_to\_election}_{ijk}$: Derived as the days remaining until the election from the poll's end date, capturing the

The model takes the form of the following equation:

$$y_{ijk} = \alpha_i + \beta_j + \gamma_k + \delta \cdot \text{pct}_{ijk} + \eta \cdot \text{sample\_size}_{ijk} + \theta \cdot \text{days\_to\_election}_{ijk} + \epsilon_{ijk} \quad (1)$$

where:

$\alpha_i \sim \mathcal{N}(\mu_\alpha, \sigma_\alpha^2)$: State-level random effect for each state i, capturing regional variations in voting patterns

$\beta_j \sim \mathcal{N}(\mu_\beta, \sigma_\beta^2)$: Pollster-level random effect for each pollster j, accounting for systematic biases or differen

$\epsilon_{ijk} \sim \mathcal{N}(0, \sigma^2)$: Error term, accounting for random noise.

γ_k : Candidate fixed effect for each candidate k, representing baseline support across states and pollsters.

δ : Coefficient for Percentage of Votes by Poll (pct_{ijk}), reflecting how poll support translates to actual vot

η : Coefficient for Sample Size (sample_size_{ijk}), weighing polls based on their reliability.

θ : Coefficient for Days to Election ($\text{days\_to\_election}_{ijk}$), capturing the trend in support as the election

3.1.1 Interpretation of Parameters

$$
\begin{aligned}
\alpha_i &: \quad \text{Captures state-specific effects, allowing the model to adjust the baseline vote share prediction based on}\\
\beta_j &: \quad \text{Accounts for systematic biases or differences in methodologies across pollsters.}\\
\gamma_k &: \quad \text{Provides an overall baseline effect for each candidate, independent of state or pollster.}\\
\delta &: \quad \text{Measures how closely polling support translates to actual vote share.}\\
\eta &: \quad \text{Adjusts the model's sensitivity to polls based on their sample size, giving more weight to larger polls.}\\
\theta &: \quad \text{Captures how support trends change as the election date approaches.}
\end{aligned}
$$

3.1.2 Prior Distributions

$$
\begin{aligned}
\alpha_i \sim \mathcal{N}(0, 2.5) &: \quad \text{State-level random effect prior for each state } i.\\
\beta_j \sim \mathcal{N}(0, 2.5) &: \quad \text{Pollster-level random effect prior for each pollster } j.\\
\gamma_k \sim \mathcal{N}(0, 2.5) &: \quad \text{Candidate fixed effect prior for each candidate } k.\\
\delta, \eta, \theta \sim \mathcal{N}(0, 1) &: \quad \text{Coefficients for polling percentage, sample size, and days to election.}\\
\sigma \sim \text{Exponential}(1) &: \quad \text{Prior for the standard deviation of the error term.}
\end{aligned}
$$

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`.

3.1.3 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 **?@tbl-modelresults**.

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…

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…

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/.