US President Prediction*

Predict the results of 2024 US President Election

Yun Chu Felix Li Wen Han Zhao

October 22, 2024

In this study, we aim to study the 2024 U.S. Presidential Election with the methodology of polls of polls across various states. With the insightful analyses on polls provided by Redfiled & Wilton Strategies, the conducted survey on the population highlights various key point for the prediction of US president. Using Generalized Linear Models (GLMs), this report analyzes the accuracy of prediction in president votes through several key variables from pollster. The model incorporates variables such as pollscore, methodology, transparency_score and hypothetical. The prediction finding highlights the significance of candicate trustworthiness among the voter decision, providing insights for swing voters. This analysis offers valuable information into the potential electoral outcomes driving the 2024 US president election.

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....

*Code and data are available at: https://github.com/younazhao/US-President-Prediction/tree/main.

2 Data

2.1 Overview

We use the statistical programming language R (R Core Team 2023)…. Our data…. 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

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 percentage of support for a candicate in a poll.

$y_i = \beta_0 + \beta_1 pollscore + \beta_2 methodology + \beta_3 transparencyscore + \beta_4 hypothetical + \epsilon$

pollscore: A numeric value representing the score or reliability of the pollster in question (e.g., -1.1). "The error and bias we can attribute to a pollster. Negative numbers are better. Stands for"Predictive Optimization of Latent skill Level in Surveys, Considering Overall Record, Empirically.

methodology: The method used to conduct the poll (e.g., Online Panel).

transparency_score: A score reflecting the pollster's transparency about their methodology (e.g., 9.0). "A grade for how transparent a pollster is, calculated based on how much information it discloses about its polls and weighted by by recency. The highest Transparency Score is 10."

hypothetical: Indicates whether the poll is about a hypothetical match-up.

We run the model in R (R Core Team 2023).

3.1.1 Model justification

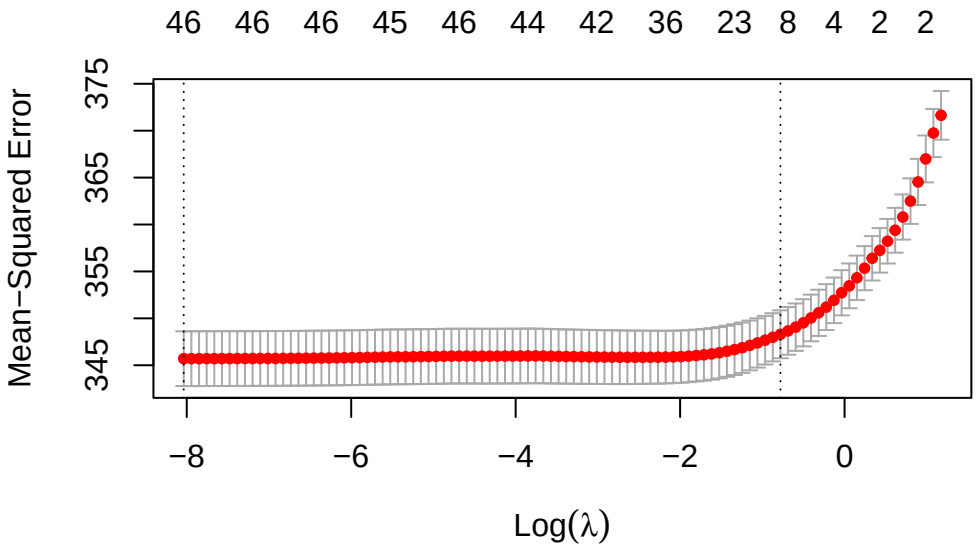
Table 1: model 1

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	63.979667	5.0205698	12.7435072	0.0000000
pollscore	-3.533768	0.3209748	-11.0094857	0.0000000
methodologyEmail	-20.892998	6.5340436	-3.1975602	0.0013902
methodologyEmail/Online Ad	-10.429453	9.0526392	-1.1520898	0.2493112
methodologyIVR	-11.034652	6.0164883	-1.8340686	0.0666730
methodologyIVR/Live Phone/Text	-13.426432	9.6688032	-1.3886343	0.1649743
methodologyIVR/Live Phone/Text/Online Panel/Email	-17.296261	5.5574331	-3.1122752	0.0018617
methodologyIVR/Online Panel	-13.633674	5.2482548	-2.5977538	0.0093971

	Estimate	Std. Error	t value	Pr(>|t|)
methodologyIVR/Online Panel/Email	-19.181827	5.1426875	-3.7299227	0.0001926
methodologyIVR/Online Panel/Text-to-Web	-19.594610	5.0543198	-3.8768045	0.0001065
methodologyIVR/Online Panel/Text-to-Web/Email	-15.913793	5.3029985	-3.0009047	0.0026983
methodologyIVR/Text	-14.497848	5.1695978	-2.8044440	0.0050499
methodologyIVR/Text-to-Web	-22.047187	5.3274275	-4.1384303	0.0000353
methodologyIVR/Text-to-Web/Email	-7.444712	9.0613151	-0.8215928	0.4113279
methodologyLive Phone	-22.009337	4.9948811	-4.4063785	0.0000106
methodologyLive Phone/Email	-19.261395	5.4088784	-3.5610700	0.0003710
methodologyLive Phone/Online Panel	-19.871909	5.1500099	-3.8586158	0.0001147
methodologyLive Phone/Online Panel/App Panel	-28.407277	7.0126693	-4.0508507	0.0000514
methodologyLive Phone/Online Panel/Text	-16.880690	5.6998052	-2.9616257	0.0030673
methodologyLive Phone/Online Panel/Text-to-Web	-14.340535	5.1969825	-2.7593964	0.0058012
methodologyLive Phone/Online Panel/Text-to-Web/Text	-14.461653	5.8205235	-2.4845966	0.0129858
methodologyLive Phone/Probability Panel	-10.465689	6.7920148	-1.5408814	0.1233767
methodologyLive Phone/Text	-3.950813	14.0219151	-0.2817599	0.7781333
methodologyLive Phone/Text-to-Web	-21.910971	5.0227875	-4.3623129	0.0000130
methodologyLive Phone/Text-to-Web/App Panel	-19.284147	6.0056340	-3.2110093	0.0013268
methodologyLive Phone/Text-to-Web/Email	-8.910865	14.0268863	-0.6352703	0.5252664
methodologyLive Phone/Text-to-Web/Email/Mail-to-Web	7.026590	6.4684401	1.0862882	0.2773772
methodologyLive Phone/Text-to-Web/Email/Mail-to-Web/Mail-to-Phone	-32.121834	7.6956175	-4.1740424	0.0000302

	Estimate	Std. Error	t value	Pr(>|t|)
methodologyLive Phone/Text-to-Web/Online Ad	-21.589862	7.6818415	-2.8105061	0.0049558
methodologyLive Phone/Text/Online Ad	-14.132632	7.4749796	-1.8906582	0.0586983
methodologyLive Phone/Text/Online Panel	-24.775149	8.2277139	-3.0111826	0.0026087
methodologyMail-to-Web/Mail-to-Phone	-19.786943	6.4005546	-3.0914419	0.0019972
methodologyOnline Ad	-25.708962	5.4662855	-4.7031869	0.0000026
methodologyOnline Panel	-17.887315	4.9695024	-3.5994179	0.0003204
methodologyOnline Panel/Email	-10.592440	7.0214561	-1.5085816	0.1314368
methodologyOnline Panel/Email/Text-to-Web	-19.288720	7.4842549	-2.5772399	0.0099732
methodologyOnline Panel/Online Ad	-10.447035	6.2383664	-1.6746427	0.0940351
methodologyOnline Panel/Probability Panel	-16.958524	7.0121219	-2.4184582	0.0156039
methodologyOnline Panel/Text	-6.538836	9.0638772	-0.7214171	0.4706695
methodologyOnline Panel/Text-to-Web	-16.626980	5.1077118	-3.2552699	0.0011365
methodologyOnline Panel/Text-to-Web/Text	-18.323110	5.1391024	-3.5654301	0.0003649
methodologyProbability Panel	-16.269851	5.0203042	-3.2408098	0.0011957
methodologyText	-13.694893	6.1228937	-2.2366701	0.0253295
methodologyText-to-Web	-24.519690	5.3181586	-4.6105602	0.0000041
methodologyText-to-Web/Online Ad	-12.714183	5.4576811	-2.3295943	0.0198471
transparency_score	-1.437809	0.0825576	-17.4158377	0.0000000
hypotheticalTRUE	-7.695244	0.4346440	-17.7047033	0.0000000

Table 2: **?(caption)**



3.2 Lasso Regularization

Lasso Regularization is performed to see if the selected variable in our model would best predict the results of polls percentage of a candidate. If the variable do not align with our actual numbers, we should consider other variables or a interaction variable.

Computing the Mean Square Error would give us a sense of our lasso model prediction.

```
Mean Squared Error: 342.4718
```

```
R-squared: 0.07872276
```

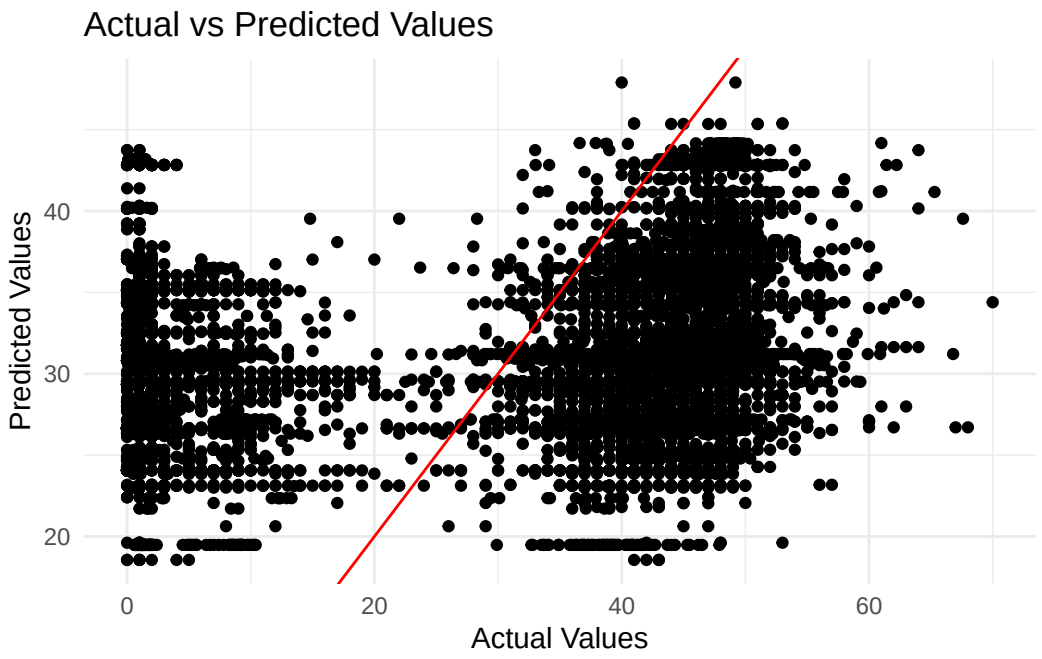
Table 3: Mean Squared Error

Metric	Value
SST	3.802858e+06
SSE	3.503487e+06
R-squared	7.872280e-02

From this graph, we can see that the cluster points do not perfectly align on the line of the best fit so we should still investigate in other variables.

Table 4: actual vs predicted

Actual	Predicted
50	43.17871
46	43.17871
48	37.02916
45	37.02916
1	37.02916
1	37.02916



Appendix

A Additional data details

A.1 pollster methodology analysis

This survey was conducted by Redfield & Wilton Strategies to assess the voting intentions of eligible voters in key U.S. swing states ahead of the 2024 Presidential Election. The primary goal of this poll is to provide an accurate and timely snapshot of public opinion in states where electoral outcomes are uncertain and could have a decisive impact on the overall result of the election. Swing states, due to their political volatility and diverse voter bases, are critical in determining the balance of power in the U.S. electoral system. Understanding voter preferences in these states is essential for political analysts, campaigns, and the general public.

A.1.1 Population of Interest

The population of interest for this survey consists of all eligible voters residing in major U.S. swing states, specifically Arizona, Florida, Georgia, Michigan, North Carolina, and Pennsylvania. These states are known for their fluctuating political alignments and are expected to play a crucial role in the upcoming election.

A.1.2 Sampling Frame

The population sampled includes eligible voters from Arizona, Florida, Georgia, Michigan, North Carolina, and Pennsylvania. Participants were selected via an online panel.

A.1.3 Sample

The sample sizes for each state were as follows:

Arizona: 750 respondents

Florida: 1,350 respondents

Georgia: 927 respondents

Michigan: 970 respondents

North Carolina: 880 respondents

Pennsylvania: 1,070 respondents

A.2 Weakness & Strength of the methodology

In terms of strengths, Redfield & Wilton has a great reputation for producing reliable polling data. They have employ a mix of online and telephone survey, which add in various resources of collecting their data. This approach can help reduce bias. In addition, Redfield & Wilton often target swing states, which makes their polls results important to the US president election.

The weakness of their methodology would incorporate a certain potential bias based on their political leaning of their clients or media. This could cause a certain neutrality in their dataset.

Overall, Redfield & Wilton has been considered as a competent, reputable and reliable pollster with his variaty and methodology on polls. Even though the pollster contained a potential bias, it has been a reliable resource in prediction of US president election.

B Model details

References

Alexander, Rohan. 2023. *Telling Stories with Data*. Chapman; Hall/CRC. https://tellingstorieswithdata.com/.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.