Prediction and analysis of 2024 US presidential election

Abstract

This paper uses the data of the 2024 American election to predict, and creates a logistics regression model to discuss the election probability of Kamala Harris. In this paper, four States are used as predictive variables to predict whether Harris is voted or not, and Michigan and Pennsylvania are more supportive of Harris as president.

Introduction

The American election, as one of the most remarkable political events in the world, not only determines the future direction of the United States, but also has a far-reaching impact on international relations and global policies. In this election, Kamala Harris and Donald Trump are the two main candidates, and their political ideas, policy propositions and personal charm will all become important factors affecting voters' decision-making. Through in-depth analysis of the 2024 presidential election in the United States, this paper aims to provide readers with a clear election picture. This paper tries to fill this gap in the academic field by collecting and analyzing the latest poll data and analyzing the characteristics of the States. Through this study, I hope readers can judge from the characteristics of states. This estimate is the difference in the voting rate of Harris in four different States.

Data

Source

The data used in this paper comes from the survey data of ABC NEWS Presidential general election polls (2024). Statistical programming language R is used to retrieve, clean and process data. In particular, the following R packages are used in this report analysis: tidyverse is used for data cleaning and processing, ggplot2 is used for data visualization, modelsummary is used for predicting the output of model data, and rstanarm is used for constructing Bayesian prediction model.

Measurement

The data set of this analysis is the data about polls from ABCnews website, which is a source with reliable data quality and good reputation. There may be measurement errors when processing the data. Small measurement errors are acceptable and have little impact on this analysis.

Examined Variables

There are many variables in the original data. This time, some variables will be selected for prediction analysis. The dependent variable of this analysis is vote, and only the candidate Kamala Harris is kept. Binary logistics regression analysis is used to predict the support rate of Kamala Harris(pct>=). This paper also selected the following variables for analysis: state. For the variable state, four states with more data are selected, namely "Georgia", "Michigan", "Pennsylvania" and "Arizona". The variable pct refers to the vote support rate, which is the numerical variables. A new variable vote is derived from this report. When pct is greater than or equal to 50%, it is defined as yes (is_vote=1), otherwise it is NO (is_vote=0).

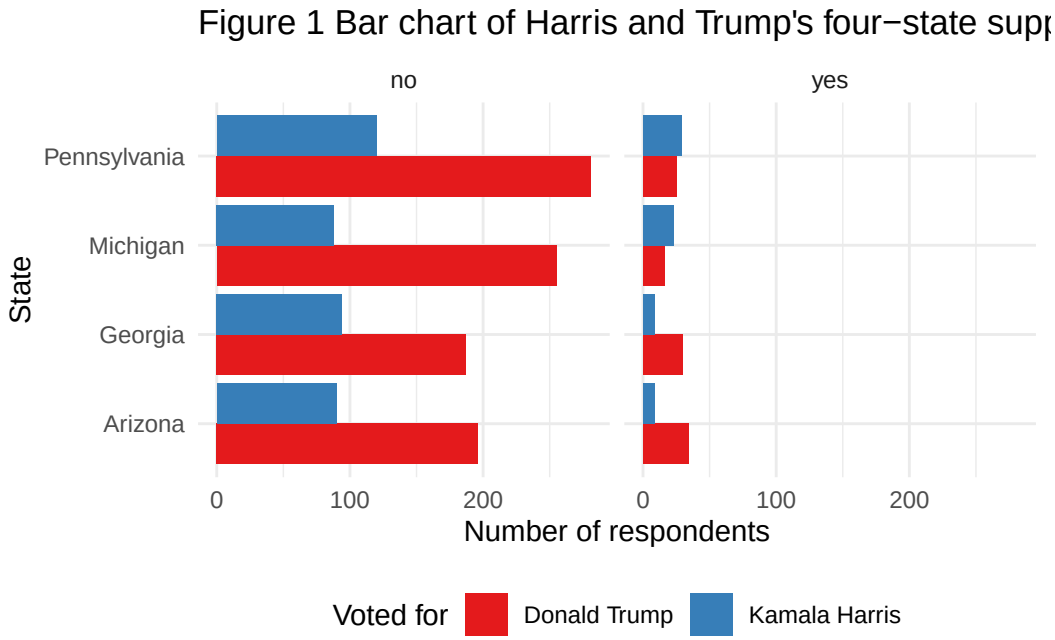

Figure 1 shows the data of American elections in four States in 2024. It can be seen that Michigan and Pennsylvania support harris more, while Georgia and Arizona support Trump more.

Model

My modeling goal is to establish a logistic regression model by using logistic distribution, because the dependent variable is a binary variable, which is used to predict the successful election rate of Harris.

Model set-up

Define yi as the whether to support Harris as president (pct>=50) and the explanatory variable as the state ("Georgia", "Michigan", "Pennsylvania" and "Arizona").

$$y_i|\pi_i \sim \text{Bern}(\pi_i)$$

$$\text{logit}(y_i) = \beta_0 + \beta_1 \times \text{state}$$

$$\beta_0 \sim \text{Normal}(0, 2.5)$$

$$\beta_1 \sim \text{Normal}(0, 2.5)$$

This paper uses the rstanarm package of Goodrich et al. (2024) to run the logistic model in R. I use rstanarm's default prior.

Model justification

According to the state data and the voting rates of Harris and Trump, this analysis shows that Michigan and Pennsylvania support Harris more, while Georgia and Arizona have lower voting rates for Harris.

Table 1: Table 1 Vote for Harris according to the state

	Support Harris
(Intercept)	−2.318
	(0.359)
stateGeorgia	−0.048
	(0.500)
stateMichigan	0.966
	(0.436)
statePennsylvania	0.887
	(0.418)
Num.Obs.	462
R2	0.027
Log.Lik.	−190.767
ELPD	−194.9
ELPD s.e.	13.5
LOOIC	389.9
LOOIC s.e.	27.0
WAIC	389.9
RMSE	0.35

Results

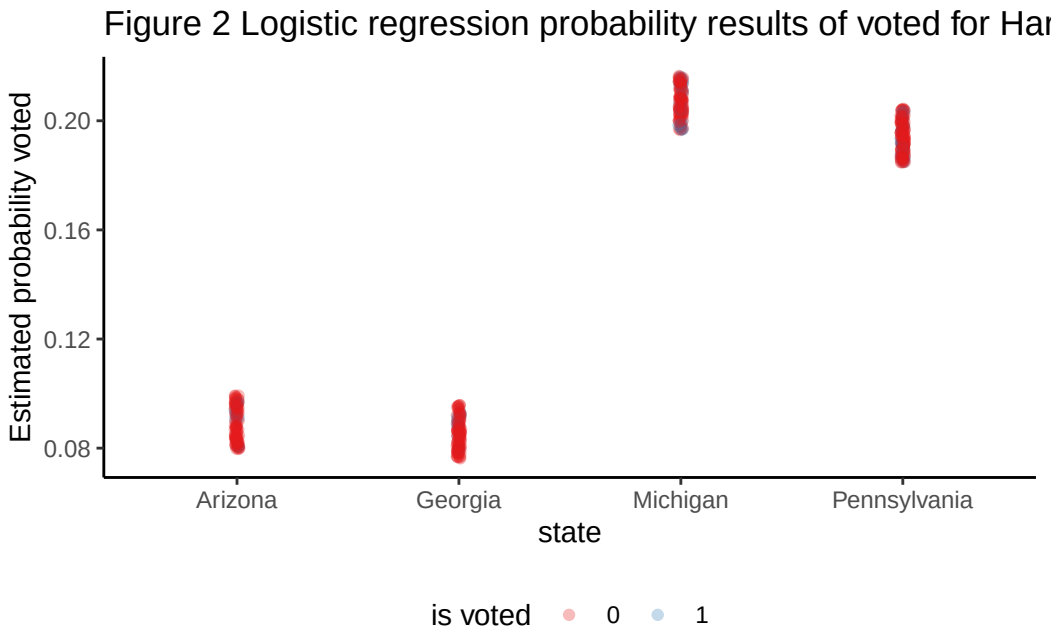

Table 1 shows that state Michigan and state Pennsylvania have significantly higher voting rates for Harris than Arizona.

Figure 2 shows that there are significant differences in the support rate of Harris among voters in different States. In Michigan, Harris' support rate is high, but it is relatively low in Arizona. These differences may be influenced by local economic conditions, social problems and the intensity of candidates' campaign activities in these States.

Discussion

This paper discusses Kamala Harris's election situation in the 2024 US presidential election, our analysis reveals that there are significant differences in the support rate of different States for her. In particular, data from Michigan and Pennsylvania show that Harris' voting rate in these areas is significantly higher than that in Arizona. This difference may be influenced by many factors, including local economic situation, social problems and the intensity of candidates' campaign activities in these States. Michigan's high support rate for Harris may reflect the state's voters' satisfaction with the current economic and social policies, and Harris's commitment to education and infrastructure investment may also attract blue-collar workers and middle-class voters in the state. Arizona's economy relies heavily on tourism and retirees, and Harris's position on these issues may have affected her support rate in the state.

Generally speaking, our analysis emphasizes that there are significant differences in Harris support rate among different States, and these differences may be affected by a series of complex factors. Through this study, we not only provide new insights for understanding the 2024 US presidential election, but also lay a foundation for future political analysis and prediction. Our research emphasizes the importance of considering state-level differences and individual-level variables when analyzing voter behavior. In the future research, I hope to mine more explanatory variables, such as the information about the respondents.

References

Arel-Bundock, Vincent. 2024. Marginaleffects: Predictions, Comparisons, Slopes, Marginal

Means, and Hypothesis Tests. https://marginaleffects.com/.

Huberfeld N ,McCuskey E ,Ulrich R M .The US Presidential Election's High Stakes for the Future of Medicine.[J].JAMA,2024,

Immigration Policies Proposed in the Major Party Platforms for the 2024 US Presidential Election[J].Population and Development Review,2024,50(3):941-947.

Long E .Last Word: Will Abortion Sway the 2024 US Presidential Election?[J].Political Insight,2024,15(3):48-48.

Morabia A .The US Presidential Election of 2024 Is a Public Health Emergency.[J].American journal of public health,2024,e1-e3.

Wickham, Hadley. 2016. Ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Appendix 1

TIPP (Trend In Politics and Policy)

TIPP focuses on the investigation and analysis of political, social and economic trends. TIPP usually uses online panels and other survey methods to collect data and provide insight into public opinions, policy changes and election dynamics. TIPP is committed to ensuring the transparency of its survey methods, including sample selection, survey design and results reporting. The survey is generally about American adult population or registered voters, and TIPP recruits respondents through online advertisements, e-mail invitations or social media platforms. TIPP usually adopts probabilistic sampling methods, such as random sampling, to ensure that every interviewee has an equal chance of being selected. Questionnaires can be distributed and collected quickly, allowing researchers to obtain a large amount of data in a short time. TIPP's online survey method allows fast, flexible and cost-effective data collection. This method can quickly adapt to the new political and social dynamics.

Appendix 2

Target population: registered voters in the United States.

Sampling method: stratified random sampling: divide the population into different layers (for example, according to geographical area, age, gender, race and political orientation), and then randomly select samples from each layer.

Sample size: according to the budget and the required error range, determine the appropriate sample size. Generally speaking, the larger the sample, the higher the reliability of the results.

Recruitment method: email invitation: use the voter registration database to send email invitations to potential interviewees. Social media advertising: Put targeted advertisements on Facebook, Twitter and other platforms to attract respondents to participate in the survey. Online market research platform: cooperate with online market research platforms such as Google Forms to recruit respondents.

Incentive measures: provide small cash rewards or gift cards to improve the response rate. Questionnaire design: design a questionnaire containing closed and open questions to ensure that the questions are concise, clear and unbiased.

Survey tools: Use online survey software (such as Google Forms and Qualtrics) to collect data.