Race to the Ballot Box: Will Demographics Decide a Winner This Year?*

A Regressional Approach to Race and Political Preferences in the 2024 US Elections and Insights on the Next

Sameeck Bhatia Sean Chua Tanmay Shinde

October 22, 2024

This paper analyzes polling data for the 2024 U.S. Presidential election to predict the winner, while also capturing demographic sentiment toward the two main contenders - Donald Trump and Kamala Harris. We found that ….[NEED TO COMPLETE]. This study highlights that considering factors such as the reputation of the pollster, quality of the poll, bias, sample size, and demographic area are essential for analyzing polling data and thereby forecasting election outcomes. Highlighting these trends is crucial for understanding the evolving political landscape in the U.S., as they reveal how voter demographics and sentiments shape election outcomes.

Table of contents

*Code and data are available at: https://github.com/SameeckBhatia/2024-US-Elections.

1 Introduction

As the 2024 U.S. Presidential election draws near, the race between former President Donald Trump and Vice President Kamala Harris has garnered widespread attention worldwide. With both candidates representing contrasting ideologies and policy priorities, appealing to distinct voter bases, understanding the public sentiment surrounding them becomes crucial in predicting the election's outcome. Polls and demographic analyses provide a lens through which we can examine the factors influencing voter preferences and gauge the political landscape as it evolves. This paper aims to analyze the 2024 election polling data to not only forecast the winner but also to capture the demographic-wise public sentiment surrounding the two contenders.

The primary estimand for this analysis is the predicted support percentage for each of the two candidates, which represents the proportion of the electorate that is expected to favor one candidate over the other. This estimand will be calculated based on survey data or polling results, and it will allow us to estimate the level of support each candidate has within the population. This analysis will also provide insights into how various demographic factors or voter preferences may influence the final outcomes.

We use polling data from FiveThirtyEight, focusing on surveys from organizations such as YouGov, Siena/NYT, and Washington Post. We used a Bayesian approach to model the polling support percentages for both candidates, and understand how polling factors such as numeric grade, poll score, state, and sample size affect the predicted support percentage.

The results show that … [TO-DO]. These results are important because … [TO-DO].

The rest of the paper is structured as follows: Section 2 details the data and measurement process. Section 3 presents the model and justifies the choices made in the building of the chosen model. Section 4 presents the results, highlighting the relationship between different variables and the polling support percentage, and Section 5 discusses the implications of the findings for this as well as future election forecasting.

2 Data

2.1 Overview

The dataset used for this paper is the 2024 national-level presidential general election dataset from FiveThirtyEight (FiveThirtyEight 2024), which compiles polling data from various reputable pollsters, including YouGov, Siena/NYT, Ipsos, and more. This dataset offers a comprehensive snapshot of public opinion leading up to the 2024 U.S. Presidential election, capturing support for the two main contenders: Donald Trump and Kamala Harris. By focusing on national polling, this dataset provides insight into voter sentiment across a wide range of demographic groups and regions.

The FiveThirtyEight dataset is part of a broader landscape of polling data used in electoral predictions. It includes surveys conducted at the national level, focusing on general election matchups, and is continuously updated as new polls are released. The dataset covers polls taken from early in the election cycle to the final days before the election, capturing the shifts in public opinion over time. While there are other polling datasets available, such as those from RealClearPolitics (RealClearPolitics 2024) or individual pollsters, we selected the FiveThirtyEight dataset due to its high level of transparency, detailed methodology grades for each pollster, and advanced polling aggregation techniques, which adjust for biases and historical pollster performance. These features provide a more refined and reliable dataset compared to others that might not incorporate such rigorous adjustments.

We use the statistical programming language R (R Core Team 2023) to perform data cleaning and analysis on the dataset to ensure consistency and reliability. Polls were filtered to include only those related to Donald Trump and Kamala Harris, with a pollster grade of 2.8 or higher, and those with a negative pollscore, ensuring reliability and unbiasedness. The dataset was further narrowed by selecting polls conducted on or after July 21, 2024, to focus on the shift in sentiment over time after the Democratic Party announced Harris as their nominee for president in the 2024 elections.

2.2 Measurement

The process of converting real-world electoral phenomena into data involves careful measurement of public sentiment toward presidential candidates, in the context of the 2024 U.S. presidential election. The dataset is built from polling data collected by various established pollsters, such as YouGov and Siena/NYT, who use a variety of survey techniques to measure voter preferences. These polls aim to capture critical variables that reflect voter sentiment, such as the percentage of respondents favoring each candidate, the sample size of the poll, and the demographics of the voters surveyed.

The measurement begins with pollsters designing surveys aimed at accurately reflecting voter behavior and sentiment. A subset of the population is sampled using stratified random sampling to ensure that the sample accurately represents the diverse demographic makeup of the electorate. Responses are collected through various channels, including phone calls, online surveys, and face-to-face interviews. These polls capture voter sentiment around a candidate as the percentage of respondents favoring the candidate.

Once the polling is done, the dataset undergoes extensive cleaning and processing to handle missing values and any discrepancies in the data. Each row in the dataset represents the outcome of a poll, including details such as the date it was conducted, the pollster's methodology, and the percentage of support for each candidate. Platforms like FiveThirtyEight then assign a numeric grade and poll score to each poll to evaluate the reliability, quality, and historical accuracy of their polling methods. The numeric grade is a standardized rating assigned to pollsters based on their historical performance, transparency, and polling methodology. The poll score measures the accuracy, consistency, and bias in each poll, pollsters that show bias toward a particular party receive a positive score.

Thus, each entry in the dataset corresponds to the results of a specific poll, providing information on when it was conducted, the pollster's methodology, and key metrics such as biasdness, sample size, and geographic coverage. By converting complex voter opinions into structured data, we are able to systematically analyze trends in public sentiment, making it possible to predict the election's outcome with greater precision.

2.3 Outcome variables

TO-DO: All variables should be thoroughly examined and explained.

TO-DO: It is important to understand what the variables look like by including graphs, and possibly tables, of all observations, along with discussion of those graphs and the other features of these data. Summary statistics should also be included, and well as any relationships between the variables.

2.4 Predictor variables

TO-DO: All variables should be thoroughly examined and explained.

TO-DO: It is important to understand what the variables look like by including graphs, and possibly tables, of all observations, along with discussion of those graphs and the other features of these data. Summary statistics should also be included, and well as any relationships between the variables.

3 Model

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam auctor quam et justo efficitur, sit amet eleifend elit euismod. Mauris non libero vel ligula tincidunt consequat. Aenean gravida, risus id auctor aliquam, orci leo auctor tellus, non tincidunt est arcu at lectus. Vivamus feugiat et quam ut consequat. Aenean mollis ullamcorper facilisis. Vivamus scelerisque lectus et elementum vulputate. Nulla et arcu vehicula, iaculis felis sit amet, semper justo. Suspendisse quam augue, hendrerit a tortor non, tristique feugiat sem. Curabitur vitae tortor nec ligula scelerisque rutrum.

3.1 Model set-up

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam auctor quam et justo efficitur, sit amet eleifend elit euismod. Mauris non libero vel ligula tincidunt consequat. Aenean gravida, risus id auctor aliquam, orci leo auctor tellus, non tincidunt est arcu at lectus. Vivamus feugiat et quam ut consequat. Aenean mollis ullamcorper facilisis. Vivamus scelerisque lectus et elementum vulputate. Nulla et arcu vehicula, iaculis felis sit amet, semper justo. Suspendisse quam augue, hendrerit a tortor non, tristique feugiat sem. Curabitur vitae tortor nec ligula scelerisque rutrum. Nullam feugiat odio metus. Cras id convallis ante, ut ornare velit. Mauris turpis purus, porttitor eu leo quis, suscipit euismod ligula.

4 Results

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam auctor quam et justo efficitur, sit amet eleifend elit euismod. Mauris non libero vel ligula tincidunt consequat. Aenean gravida, risus id auctor aliquam, orci leo auctor tellus, non tincidunt est arcu at lectus. Vivamus feugiat et quam ut consequat. Aenean mollis ullamcorper facilisis. Vivamus scelerisque lectus et elementum vulputate. Nulla et arcu vehicula, iaculis felis sit amet, semper justo. Suspendisse quam augue, hendrerit a tortor non, tristique feugiat sem. Curabitur vitae tortor nec ligula scelerisque rutrum. Nullam feugiat odio metus. Cras id convallis ante, ut ornare velit. Mauris turpis purus, porttitor eu leo quis, suscipit euismod ligula.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam auctor quam et justo efficitur, sit amet eleifend elit euismod. Mauris non libero vel ligula tincidunt consequat. Aenean gravida, risus id auctor aliquam, orci leo auctor tellus, non tincidunt est arcu at lectus. Vivamus feugiat et quam ut consequat. Aenean mollis ullamcorper facilisis. Vivamus scelerisque lectus et elementum vulputate. Nulla et arcu vehicula, iaculis felis sit amet, semper justo. Suspendisse quam augue, hendrerit a tortor non, tristique feugiat sem. Curabitur vitae tortor nec ligula scelerisque rutrum. Nullam feugiat odio metus. Cras id convallis ante, ut ornare velit. Mauris turpis purus, porttitor eu leo quis, suscipit euismod ligula.

5 Discussion

5.1 First discussion point

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam auctor quam et justo efficitur, sit amet eleifend elit euismod. Mauris non libero vel ligula tincidunt consequat. Aenean gravida, risus id auctor aliquam, orci leo auctor tellus, non tincidunt est arcu at lectus. Vivamus feugiat et quam ut consequat. Aenean mollis ullamcorper facilisis. Vivamus scelerisque lectus et elementum vulputate. Nulla et arcu vehicula, iaculis felis sit amet, semper justo. Suspendisse quam augue, hendrerit a tortor non, tristique feugiat sem. Curabitur vitae tortor nec ligula scelerisque rutrum. Nullam feugiat odio metus. Cras id convallis ante, ut ornare velit. Mauris turpis purus, porttitor eu leo quis, suscipit euismod ligula.

5.2 Second discussion point

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam auctor quam et justo efficitur, sit amet eleifend elit euismod. Mauris non libero vel ligula tincidunt consequat. Aenean gravida, risus id auctor aliquam, orci leo auctor tellus, non tincidunt est arcu at lectus. Vivamus feugiat et quam ut consequat. Aenean mollis ullamcorper facilisis. Vivamus scelerisque lectus et elementum vulputate. Nulla et arcu vehicula, iaculis felis sit amet, semper justo. Suspendisse quam augue, hendrerit a tortor non, tristique feugiat sem. Curabitur vitae tortor nec ligula scelerisque rutrum. Nullam feugiat odio metus. Cras id convallis ante, ut ornare velit. Mauris turpis purus, porttitor eu leo quis, suscipit euismod ligula.

5.3 Third discussion point

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam auctor quam et justo efficitur, sit amet eleifend elit euismod. Mauris non libero vel ligula tincidunt consequat. Aenean gravida, risus id auctor aliquam, orci leo auctor tellus, non tincidunt est arcu at lectus. Vivamus feugiat et quam ut consequat. Aenean mollis ullamcorper facilisis. Vivamus scelerisque lectus et elementum vulputate. Nulla et arcu vehicula, iaculis felis sit amet, semper justo. Suspendisse quam augue, hendrerit a tortor non, tristique feugiat sem. Curabitur vitae tortor nec ligula scelerisque rutrum. Nullam feugiat odio metus. Cras id convallis ante, ut ornare velit. Mauris turpis purus, porttitor eu leo quis, suscipit euismod ligula.

5.4 Weaknesses and next steps

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam auctor quam et justo efficitur, sit amet eleifend elit euismod. Mauris non libero vel ligula tincidunt consequat. Aenean

gravida, risus id auctor aliquam, orci leo auctor tellus, non tincidunt est arcu at lectus. Vivamus feugiat et quam ut consequat. Aenean mollis ullamcorper facilisis. Vivamus scelerisque lectus et elementum vulputate. Nulla et arcu vehicula, iaculis felis sit amet, semper justo. Suspendisse quam augue, hendrerit a tortor non, tristique feugiat sem. Curabitur vitae tortor nec ligula scelerisque rutrum. Nullam feugiat odio metus. Cras id convallis ante, ut ornare velit. Mauris turpis purus, porttitor eu leo quis, suscipit euismod ligula.

Appendix

References

FiveThirtyEight. 2024. *FiveThirtyEight: 2024 US Presidential Election Polls*. https://projects.fivethirtyeight.com/polls/president-general/2024/national/.

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

RealClearPolitics. 2024. "RealClearPolitics Polling Data." https://www.realclearpolitics.com/epolls/latest_polls/.