Assuming Everyone Voted, Trudeau Would Win 2019 Canadian Federal Election with 34% of Popular Vote*

Xinyi Zhang

13 December 2020

Abstract

In the 2019 Canadian Federal Election, incumbent Prime Minister Justin Trudeau claimed a narrow victory over the Conversative Party amid 62% voter turn out. In this paper, we develop a multi-level regression model with post-stratification by training a logistic regression model using voter survey results and predicting the outcome of the popular vote using large-scale demographic data for the Canadian population. Our model predicts that assuming 100% voter turnout, Trudeau would have won the 2020 Canadian Federal Election with a larger margin. Our prediction speaks to the importance of encouraging voter turnout and our breakdown of votes by demographic group could potentially provide the candidates of the election with information on how to target voters.

Keywords: forecasting, 2019 Canadian Federal Election, Trudeau, Andrew Scheer, multilevel regression with post-stratification

1 Introduction

In 2019, Trudeau defied election predictions and narrowly won enough seats to win reelection. This raises the question, what would the election result have been if everyone voted? By choosing 5 explanatory demographic variables closely associated with voting and political affiliation (age, gender, state, race, and education), we develop a multi-level regression model with post-stratification in order to predict the outcome of the popular vote.

In our analysis, we modeled the relationship between our selected demographic variables and a person's likelihood to vote for either Trudeau or Scheer. We analyzed the significance of our model, and the importance of using multilevel regression with post-stratification because the training data is not proportional to the American population. We found that we were able to make predictions with approximately 70% accuracy and the strongest indicators of 2019 voting were education and gender.

This paper discusses the 2 datasets we used, how they were collected and key highlights of these datasets, followed by visualizations of the data. Next we explained the construction of our model and the positives and negatives of extrapolating information from a smaller voter survey to the Canadian population using a post-stratification dataset. Finally, we present our results and discuss how our results should inform future political strategies.

*Code and data supporting this analysis are available at: https://github.com/cindyzhang99/sta304_ps5.

2 Data

3 Model

4 Results

5 Discussion

6 References

http://www.ces-eec.ca/ https://doi.org/10.7910/DVN/DUS88V

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Xie, Yihui. 2020. *Bookdown: Authoring Books and Technical Documents with R Markdown*. https://github.com/rstudio/bookdown.