Impact of Voter Turnout on General Elections: A Statistical Analysis on the 2019 Canadian Federal Election

James Bai

12/22/2020

Abstract

Endless studies can be done to analyze voter turnout and election results. This report analyzes what would have happened if everyone voted in the 2019 Canadian Federal Election, as the voters turnout rate was only around 67%. Using logistic regression with post-stratification, the estimates suggest that the Liberal Party would have still edged over the Conservative Party in the election.

Keywords

logistic regression, post-stratification, election, liberal party, conservative party, voter turnout.

Introduction

Voter turnout is a popular discussion among political science experts. As the number of democracies rise in the world, the amount of participation during elections has always been studied. Additional research is needed to study the impact voter turnout has on election results in Canada. What would have happened if every Canadian voted in the 2019 Canadian Federal Election? Would Justin Trudeau have won the election? Or would Andrew Scheer pull one over Trudeau?

Despite the Liberal Party winning the 2019 Canadian Federal Election, the Conservative Party actually pulled in more votes. However, due to the first-past-the-post system, the Liberal Party edged out a minority government. Considering only 67% of eligible voters came out to vote, the election results may have been different if everyone voted (Election Canada, n.d.).

Previous research found that increased voter turnout does not have an effect on socialist parties. The article by Schaefer (2012) uses an empirical study on 1,500 urban areas in Germany and found that city-centres that experienced the largest drop in turnout are also more heavily leaned towards socialist parties. That research is focused on the impact turnout has on more socialist parties and is missing a viewpoint on other party ideologies. . Therefore, this study will examine Canadian voters using their census data and determine if their voting habits may have had an impact on the election results.

More specifically, this study will be using Multilevel Regression with Post-Stratification using the 2019 Canadian Election Study to predict who would win the election if everyone voted in the election.

Methodology

Data

The survey data was from the 2019 Canadian Election Study. Some cleaning was done upon the survey data, the key variables kept from the survey were the age range, gender, whether they would vote for the liberal or not, and whether they would vote for the conservative or not. Observations who were not a Canadian Citizen were removed from the the study as they do not have to right to vote in the Canadian Federal Election. The census data used in this study was web scraped from Statistic Canada. The census data consist of the population count for each age group and gender as they are the predictors used for this study.

Model

The two models used in this study are two logistic regression model to model the percentage of voters who would vote for the Liberal Party and Conservative Party. The predictors used in this study will be age group and gender as this is a post election anlyasis to study whether Justin Trudeau deserves his victory if the entire population voted in the election, so not many predictors would be needed. The logistic regression model for the Liberal Party would be:

$$log(\frac{\hat{p}_{Liberal}}{1-\hat{p}_{Liberal}}) = \beta_{L0} + \beta_{L1}x_{age\_group} + \beta_{L2}x_{gender}$$

and the model for the Conservative Party would be:

$$log(\frac{\hat{p}_{Conservative}}{1-\hat{p}_{Conservative}}) = \beta_{C0} + \beta_{C1}x_{age\_group} + \beta_{C2}x_{gender}$$

Where $\hat{p}_{Liberal}$ and $\hat{p}_{Conservative}$ are the proportion of voters from the survey who would vote for for the Liberal and Conservative Party respectively, B_{L0} and B_{C0} are the intercept, B_{L1} and B_{C2} are the coefficient for age group, and B_{L2} and B_{C2} are the coefficient for gender.

Post-Stratification

Using Post-stratification analysis, an estimate that represents the entire Canadian population can be achieved. Post-Stratification analysis partition the data into many cells, then estimates response variables within each cells. Then a population level estimate would be achieved by dividing the sum of the population proportion of each cell to the total population.

The general formula is

$$\hat{y}^{PS} = \frac{\sum N_j \hat{y}_i}{\sum N_j}$$

Result

The estimate of the proportion of voters who would vote for the Liberal Party is $\hat{y}^{PS}_{Liberal} = 0.2858$ and the proportion of voters who would vote for the Conservative Party is $\hat{y}^{PS}_{Conservative} = 0.2818$. Comparing the two estimate, the chances of Trudeau's party winning is slightly above Scheer's.

While the estimate for both parties may seem low, there were around 5000 of the 31000 observation from the survey didn't indicate who they would vote for. If those observation were removed from the survey, the new estimate of the proportion of voters who would vote for the Liberal Party is $\hat{y}^{PS}_{Liberal} = 0.3357$ and the

estimate for the Conservative Party would be $\hat{y}^{PS}_{Conservative} = 0.3304$. This is a really good result from the model, as the final vote share in the actual election is 33.1% and 34.4% of the voters respectively.

Based off the model, the estimates suggest that the Liberal Party would slightly beat the conservative party in the votes share. Even though the estimate of the vote share of both parties does not even add up to 70% of the final votes, in the actual election, a lot of the other 33% of the votes ended up being divided by Bloc Québécois, New Democratic Party, and the Green Party. None of them had a real chance in competing against the the Liberal and Conservative Party.

Discussion

Summary

In this study, two post-stratification on two separate logistic models were used to find the estimate of the proportion of voters that would vote for Liberal and Conservative party respectively.

Conclusions

In conclusion, Justin Trudeau most likely would have won the election by minority even if the entire Canadian population voted in the 2019 Canadian Federal Election. The estimate even suggest that the Liberal Party would be getting a higher vote share than the Conservative Party as opposed to the actual result.

Weakness & Next Steps:

One of the weaknesses of this report is that the actual election is not won by having most votes, but rather by having the most seats in the House of Common. One way to improve in the future, using post-stratification for each individual riding, then compare which party would have the most seats in the House of Common.

References

Alexander, R, Caetano, S (2020) 01-data_cleaning-survey [R]. q.utoronto.ca

Caetano, S (2020) ProblemSet3-template [R]. q.utoronto.ca

CBC News Labs. (n.d.). Canada Votes 2019. CBC News. https://newsinteractives.cbc.ca/elections/federal/2019/results/

Election Canada. (n.d.). Voter turnout at federal elections and referendums. https://www.elections.ca/content.aspx?section=ele&dir=turn&document=index&lang=e

Schaefer, A. (2012). Does declining voter turnout affect election results? An analysis of election data from districts in major German cities. Politische Vierteljahresschrift, 53(2), 240–264.

Statistics Canada. (2017). Census profile, 2016 census (98-316-X2016001) [Data file]. Statistics Canada. https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/page.cfm?La

Stephenson, Laura B; Harell, Allison; Rubenson, Daniel; Loewen, Peter John. (2020) "2019 Canadian Election Study - Online Survey", https://doi.org/10.7910/DVN/DUS88V, Harvard Dataverse, V1

Github Repo

https://github.com/jamesbyebai/Final-Project