

Analysis of the 2019 Canadian Federal Election If Everyone Voted

Eric Yuan

December 21th, 2020

GitHub Repository: <https://github.com/ewyuan/2019-canadian-federal-election-analysis>

Abstract

Low voter turnout at elections in a democracy is problematic, however, how would an election's results be different if the voter turnout rate was 100%? The purpose of this analysis is to determine what the 2019 Canadian Federal Election results would look like if every eligible voter had voted. Using multilevel regression with post-stratification (MRP) we determined that the Liberal Party received 30.4% of votes when the voter turnout was 100%.

Keywords

Election, Voter Turnout, Logistic Regression, Post-stratification, Multilevel regression with post-stratification

Introduction (Section 1)

In a democratic country, voting is considered to one of the most important rights the people have; voting allows for citizens of a democratic nation to vote for the candidate that best aligns with their values. For the 2019 Canadian Federal Election, the voter turnout was 67.0% [1], with the Liberal Party winning 157 seats out of 338 seats in the House of Commons, however losing the popular vote to the Conservative Party, forming a minority government. Had the voter turnout be 100.0%, could the Liberal Party have won the popular vote? This paper aims to analyze what the election results would look like in the 2019 Canadian Federal Election had everyone who was legally allowed to vote, voted.

To determine what the outcome of the 2019 Canadian Federal Election would like if the voter turnout rate was 100.0%, using multilevel regression with post-stratification (MRP). We are going to perform logistic regression with 2019 Canadian Election Survey dataset [2][3], and then post-stratify with 2016 Census Public Use Microdata Files which contains individualized characteristics of the Canadian population [4].

In Section 2, we will be going over the methodology of our analysis; primarily, we will be going over datasets that we used and the variables we chose to train our logistic regression, but also go in-depth into the post-stratification process. Lastly, we will give our results in Section 3, and discuss the results in Section 4.

Methodology (Section 2)

Data

The data that was used to train our logistic model was obtained from the Canadian Election Survey (CES)[2][3]. This dataset had responses from 37,822 surveyees regarding general sentiment on 2019 Canadian Federal Election, as well as who the surveyees planned on voting for in the 2019 Federal Election. The CES dataset had over 600 variables, however we only focused on six: gender, education, total income, province of residence, age and who they planned on voting for in the 2019 Canadian Federal Election; given that voters in Canada have the option to vote for multiple parties (NDP, Conservative Party, Liberal Party, Bloc Quebecois, Green, People’s, and Other), we created another column for whether or not they planned on voting for Trudeau (from the Liberal Party), with binary values 0 and 1, where 0 represents not voting more Trudeau and 1 represents voting for Trudeau. Furthermore, all observations with “Don’t know” or NA values were removed from the dataset.

The other dataset that we used for post-stratification was the 2016 Canadian Census [4]. From the 2016 Canadian Census, we obtained over 600,000 individual census respondents, and isolated the same variables we used to train our model: gender, education, total income, province of residence and age. Furthermore, all observations with “Don’t know” or NA values were removed from the dataset.

Model

To model the proportion of all eligible voters in Canada that would vote for Trudeau, a logistic model will be used. The final logistic model had **age**, **gender**, **income**, **education** and **province** as the predictors to predict the response variable, represented by the equation:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \sum_{i=1}^K \beta_i x_i$$

where the response variable p represents whether or not they planned on voting for Justin Trudeau (0 indicates not voting for Trudeau, 1 indicates voting for Trudeau), β_0 represents the intercept term which is the initial starting point in the case all the predictors are 0, K represents the number of predictors, β_i indicates the slope of the predictor, and x_i indicates the value of the predictor.

Post-Stratification

The post-stratification technique was used to create a better representation of the Canadian voting population. With post-stratification, we assume that the demographic information is collected, each observation in the dataset is categorized into specific cells by the demographic variables. Each estimation for the variable of interest within each cell, can construct a population level estimate by assigning weights to each cell, relative to its proportion in the 2016 Census Public Use Microdata Files dataset. The formula that we use for the post-stratification technique is:

$$\hat{y}^{PS} = \frac{\sum_{j=1}^K N_j \hat{y}_j}{\sum_{j=1}^K N_j}$$

where K is the number of cells, \hat{y}_j is the estimate proportion for a given cell, and N_j is the size of the cell.

In the analysis completed in this report, the 2016 Census Public Use Microdata Files dataset was split into cells by education, age, income, province and gender. The reason we chose these predictors is because these common indicators that strongly influence who an individual would vote for [6]. This resulted into 9,874 cells that were created.

Results (Section 3)

From Table 1, we can see that geographic residence has the largest impact on whether or not someone will vote for Justin Trudeau in our model. If a voter is from Newfoundland and Labrador, the log odds of them voting for Trudeau increases by a factor of 1.35.

term	estimate	std.error	statistic	p.value
as.factor(province)Newfoundland and Labrador	1.3472887	0.2272909	5.927596	0.0000000
as.factor(province)Nova Scotia	1.2591155	0.1878565	6.702538	0.0000000
as.factor(province)Ontario	1.1398956	0.1083986	10.515777	0.0000000
as.factor(province)New Brunswick	1.1284194	0.2179406	5.177646	0.0000002
as.factor(province)Quebec	1.0902232	0.1176567	9.266142	0.0000000
as.factor(province)Prince Edward Island	0.8699938	0.3895799	2.233159	0.0255385

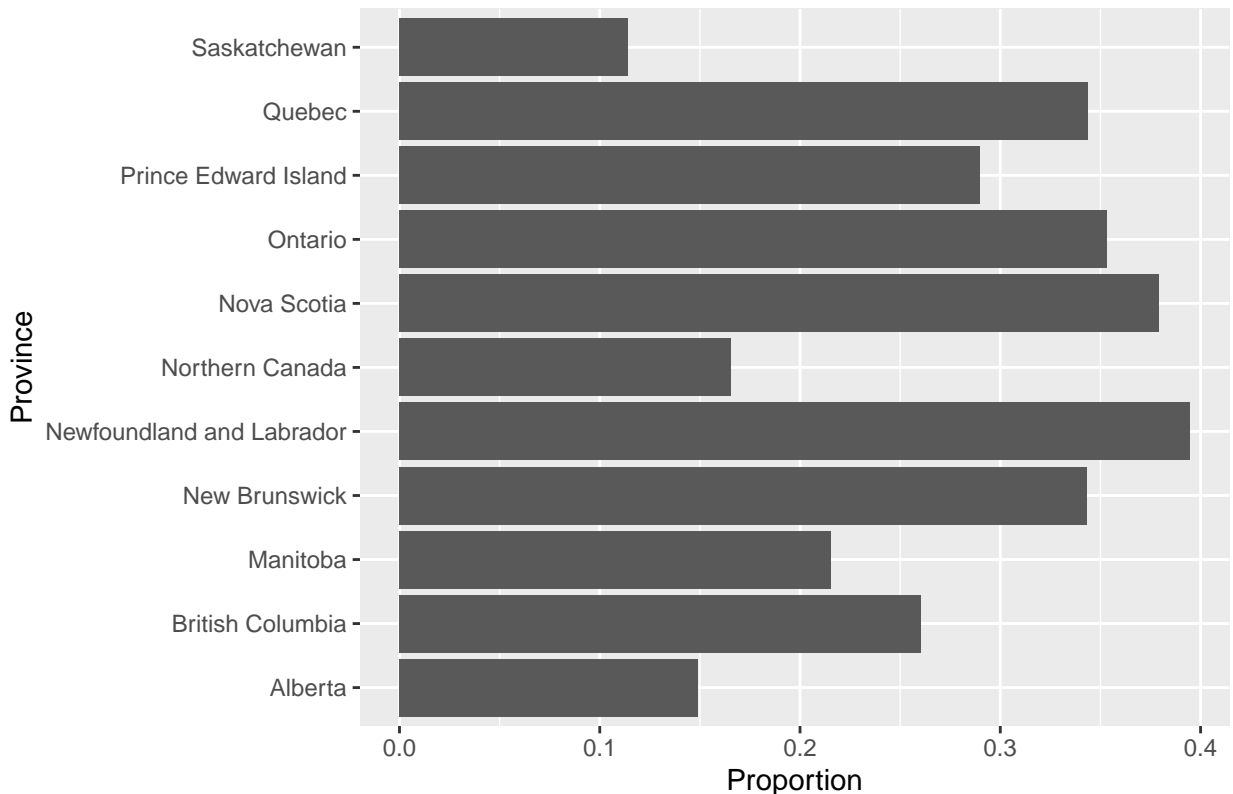
(Table 1: Top 6 Model Estimates)

From using the post-stratification technique, we can see that if all eligible voters voted in Canada in the 2019 Federal Election, the Liberal Party would have received 30.4% of the votes.

```
## [1] 0.3042117
```

In addition to the estimates from Table 1, we can see in Plot 1 that the same provinces listed in Table 1 have the highest proportion of voting for the Liberal Party, with Newfoundland and Labrador at 0.395, Nova Scotia at 0.379, Ontario at 0.353, New Brunswick at 0.343, Quebec at 0.344, and Prince Edward Island at 0.290.

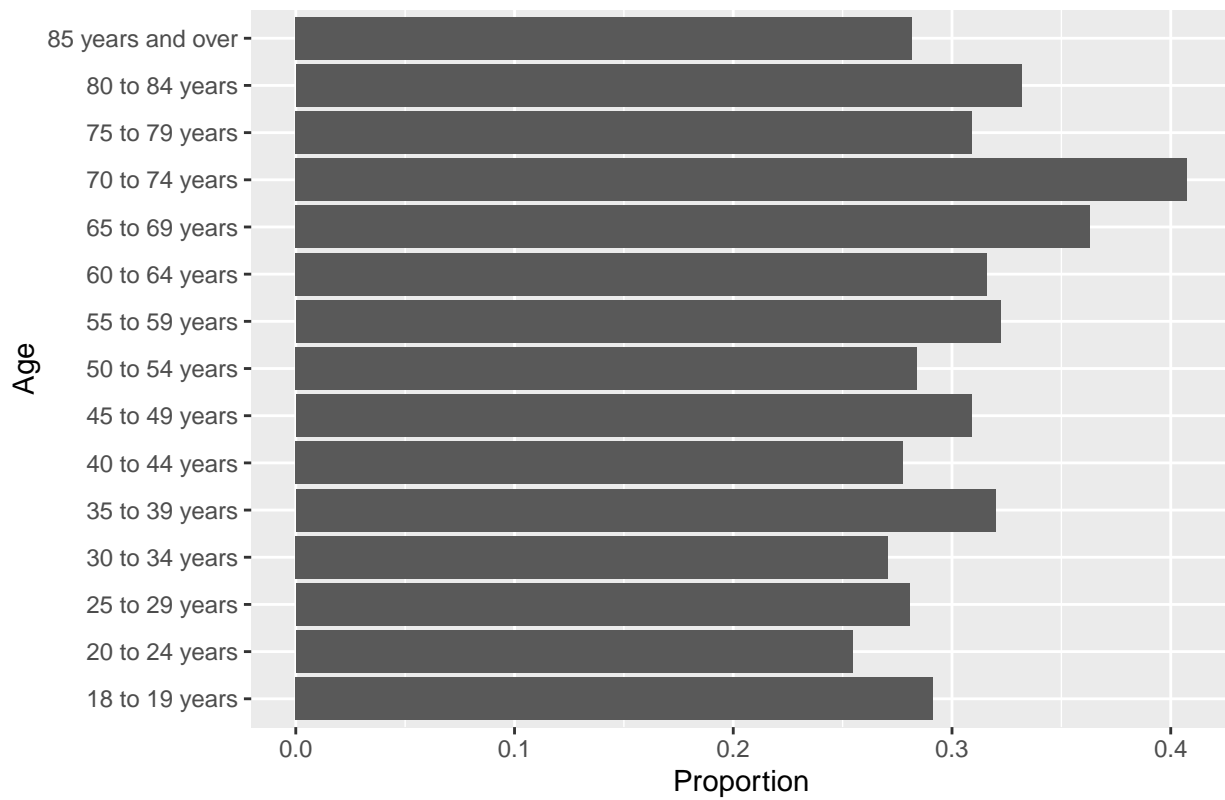
Proportion of voters that would vote for Trudeau, grouped by province



(Plot 1: Proportion of voters that would vote for Trudeau, grouped by province)

We also found in Plot 2 that individuals that are 55 years old to 84 years old have a higher proportion of the sub-population voting for the Liberal Party.

Proportion of voters that would vote for Trudeau, grouped by age



(Plot 2: Proportion of voters that would vote for Trudeau, grouped by age)

Discussion (Section 4)

Summary

We used a logistic model to model what the election results would look like if everyone who was eligible to vote in the 2019 Canadian Federal Election voted. Specifically, what would the election results look like for the Liberal Party if every eligible voter voted? Using data from the Canadian Election Survey, we selected the predictor variables: age, income, gender, residing province, and education to predict if they would vote for the Liberal Party. A post-stratification technique was then used by partitioning the data retrieved from the 2016 Individualized Canadian Census into their own cells to create a better representation of the Canadian voting population.

Conclusion

From our results, we can see that the Liberal Party will receive 30.4% of all votes if every eligible voter had voted. The actual results of the 2019 Canadian Federal Election resulted in the Liberal Party winning 157 seats, and 33.1% of the votes: forming a minority government [5]. Since the proportion of votes for the Liberal Party is not significant enough guarantee the number of seats to form a minority or majority government, we cannot conclude that the Liberal Party will be able to form a government.

Weaknesses

There are two main weaknesses of our analysis: we only consider if voters will vote for the Liberal Party, and we fail to look at the Federal electoral districts. Unlike the United States of America, Canada has more than 2 main parties, so using a logistic model to model election results does not give us enough information on which party has the most votes. Moreover, the analysis is done by looking at provinces, rather than Federal electoral districts, because Canada uses a first-past-the-post system, the candidate that has the votes wins the electoral district, and secures a seat for their party. Another weakness is that our model only looks at five parameters.

Next Steps

To improve upon the weaknesses of our analysis, we can use a multinomial logistic model, and perform the analysis on Federal electoral districts instead of the province as a whole. By using a multinomial logistic model, we will be able to obtain the proportion of voters every party in the electoral district will receive; this information is extremely valuable as it will allow us to determine the elected official of electoral districts, and also help us keep track of how many seats each party received. These two improvements give another level of insight that will allow us to clearly determine which party will be able to form a government, but also how many votes they received if every eligible voter had voted in the 2019 Canadian Federal Election.

Furthermore, we can also add additional predictors, such as geologically specific variables that give a better idea of what the political environment of an electoral district is, such as what the current elected party of the electoral district is.

References

- [1] <https://www.elections.ca/content.aspx?section=ele&dir=turn&document=index&lang=e>
- [2] 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
- [3] Stephenson, Laura, Allison Harrel, Daniel Rubenson and Peter Loewen. Forthcoming. 'Measuring Preferences and Behaviour in the 2019 Canadian Election Study,' Canadian Journal of Political Science.
- [4] <https://www150.statcan.gc.ca/n1/en/catalogue/98M0001X>
- [5] <https://www.cpac.ca/en/cpac-in-focus/vote-2019-live-results/>
- [6] <https://www150.statcan.gc.ca/n1/pub/75-001-x/2012001/article/11629-eng.htm#a1>