Estimation of 2019 Canadian Election results given every voter participated

Min Zhang

22 December 2020

Abstract:

Canadians made their choices to elect a minority government led by Justin Trudeau in 2019. However, this minority government lost its popular vote as narrow as 1% to the Conservatives. In this paper, we strive to answer one question, will the Liberals win the popular vote if every eligible voter shows up in the poll? In service of this goal, we implement six individual logistic regression models with post-stratification technique. As a result, the Conservatives would win an even bigger margin against the Liberals if every eligible voter counted. Code and data supporting this analysis are available at https://github.com/charleszhangsg/STA304-Final-Project.git

Keywords:

2019 Canadian Gederal Election, Liberals, Conservatives, popular vote, logistic regression, post-stratification, constituency

Introduction:

The 43rd Canadian general election was held on 23 October 2019, which aimed to elect members of the House of Commons. The result shows a slight fall in voter turnout, as "roughly 66 per cent of eligible voters cast ballots" (Hayes, 2019) compared with "68.3 per cent of eligible voters in 2015" (Hayes, 2019). Justin Trudeau's Liberal Party of Canada lost its majority but secured most electoral seats of 157 in the House of Commons. In contrast, the Liberal lost the popular vote to the Conservatives who won the "34.34% of popular votes" ("Election 2019 Results" 2019).

The relatively low participation in politics led us to think whether the Liberals would win the most popular vote against the Conservatives if every eligible Canadian voter shows up in the poll, as "young people contributed mightily to that big increase in voter participation"(Grenier 2019) and that "those voters played an important role in the Liberal victory"(Grenier 2019).

To predict our assumption, we select two datasets - 2017 General Social Survey on Family (2017 GSS) and Canada Election Study 2019 phone survey (2019 CES) as our census data and survey data, respectively. The 2017 GSS is a sample survey conducted from February 2

to November 30 2018 via telephone and its target population includes all Canadians who are 15 or over living in 10 provinces. The CES2019 "provides insight into such topics as the intentions of voters, what issues voters deem important, and the perception of parties and candidates" (Canadian Election Study, 2019).

In consequence, we select variables gender, age and education from these two datasets and implement a multilevel regression with post-stratification technique to predict if the Liberals could win the most popular votes in the 2019 Canadian general election given every vote counted. This paper will also include the next step we could take to make the analysis more accurate and convinced.

Methodology:

In this section, we will introduce two datasets that we converted as census data and survey data for the multilevel regression with post-stratification analysis. We will also explain how we approach the raw data and the procedures on how we manipute. This section will also inclde a multiple regression model applied to six major political parties in 2019 Canadian general election to estimate their individual portion of the popular vote.

Data:

census data:

The dataset selected for the census data is the 2017 General Social Survey(GSS) in which it collectd social data between February and November of 2017 by Statistics Canada. The goal of this survey is to "monitor changes in the living conditions and well-being of Canadians over time, and to provide information on specific social policy issues of current or emerging interest" (Statistics Canada 2019).

In order to make the analysis more legimate, we exclude those cases below the age of 18 because a voter must be a Canadian citizen aged 18 or older on election day. The variables we are interested are gender, age, province and education. We include these predictors because we want to investigate if a voter's political choice is associated with his/her age and educational background, as people in different age groups tend to have different political preferences which may affect their chocie in the ballot. The level of education would also alter individual's political preference, because people who have higher educational background tend to have different political view and interest than those people who did not achieve higher education.

The exception is in this datatset we include the predictor 'province' which indicates a specific province/territory a voter lives. The multilevel regression analysis will not include this predictor for research purpose, however, this predictor is useful when we are discussing about the next step. We also manipulate the predictor called education to match the same categories that the survey data demonstrates.

survey data:

The survey data we choose from is the 2019 Canadian Election Study (2019 CES) which provides topics like intentions of voters and perceptions of political parties in Canada. We select the same predictors as in the census data and manipulate predictors' names and values to be corresponding to those in the census data too. For example, we changed "q3" to be "gender", because "q3" is a question in which it asks the gender of the respondent.

We also exclude the voting choices other than six major political parties in Canada, because six parties account for over 99% of popular vote. Besides, in order to counter every eligible voter, we also remove those cases identified as "None of these", "Will not vote", "Will spoil ballet", "Refused" and "Don't know/undecided", because manipulating these values or randomly assigning them will dramatically affect the election result. As such, we consider the remaining votes as "good" votes and count every single of them in the analysis.

Model:

Our estimation in popular vote relies on the logistic regression model for each single party. In the situation of 2019 Canadian general election, we choose six major political parites who account for 99.26% of popular votes - The Liberal Party of Canada, The Conservative Party of Canada, New Democratic Party(NDP), Bloc Québécois, Green Party and People's Party in the analysis. Each individual party will receive a predicted result in the analysis.

Table 1: Model Breakdown

Model	Outcome Variable	Reference Level	Description
Model 1	Liberal_popular_vote	vote for other five parties	In this model, we exclude undecided voters, voters who will not vote, and voters who vote not for the Liberals.
Model 2	Conservative_popular_vote	vote for other five parties	In this model, we exclude undecided voters, voters who will not vote, and voters who vote not for the Conservatives.
Model 3	NDP_popular_vote	vote for other five parties	In this model, we exclude undecided voters, voters who will not vote, and voters who vote not for the NDP.
Model 4	Bloc_Québécois_popular_vote	vote for other five parties	In this model, we exclude undecided voters, voters who will not vote, and voters who vote not for the Bloc Québécois.
Model 5	Green_popular_vote	vote for other five parties	In this model, we exclude undecided voters, voters who will not vote, and voters who vote not for the Green Party.

Model 6	People_popular_vote	vote for other five parties	In this model, we exclude undecided voters, voters who will not vote, and voters who vote not for the People's Party.

Model Specifics:

Each model for the individual party will take the same mathematical form as below because we will use the same variables in each model to predict the proportion of the popular vote each political party could gain in the election. Table 2 will show a more specific model breakdown.

$$\log\left(\frac{\hat{p}}{1-\hat{p}}\right) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \underbrace{(\hat{\beta}_3 x_{3,1} + \cdots + \hat{\beta}_7 x_{3,5})}_{\text{education categorical variables}}$$

Table 2: Breakdown of Variable Categories

Notation	Variable	Category	Reference Level
x_1	age	Not Applicable - Continuous Variable	Not Applicable - Continuous Variable
x_2	gender	Male	Female
$x_{3,1}$	education	High school diploma or a high school equivalency certificate	Less than high school diploma or its equivalent
$x_{3,2}$	education	College, CEGEP or other non-university certificate or diploma	Less than high school diploma or its equivalent
$x_{3,3}$	education	University certificate or diploma below the bachelor level	Less than high school diploma or its equivalent
$x_{3,4}$	education	Bachelor degree (e.g. B.A., B.Sc., LL.B.)	Less than high school diploma or its equivalent
$x_{3,5}$	education	University certificate, diploma or degree above the bachelor level	Less than high school diploma or its equivalent

The left side of the equation is the log-odds of a binary result, where p is the probability of our event of interest happening (Caetano, 2020a). In model 1, the left side represents the log-odds of the percentage of popular vote that the Liberal Party of Canada would win. Similarly, we can interpret this expression from model 2 to model 6.

The slope coefficients $\hat{\beta}_1, \ldots, \hat{\beta}_7$ represent the change in log-odds attributed to its associated predictor variable $x_1, \ldots, x_{3,5}$ (Caetano 2020a). We include three predictor variables in each model, including age, gender and education. Therefore, in Model 1, $\hat{\beta}_1$ represents the change in log-odds that an individual aged x_1 will vote for the Liberals. Likewise, in Model 2, $\hat{\beta}_2$ indicates the change in log-odds that female- relative to a male-identifying individual will vote for the Conservatives. Finally, $\hat{\beta}_0$ is the intercept of our logistic regression model. It represents the log-odds of our binary result outcome variable, given that all of our

predictor variables are zero. For example, in Model 1, it represents the log-odds that an individual who is eighteen, male and whose education is less than high school will vote for the Liberals.

Post-stratification:

Like we mentioned in the introduction, we will implement post-stratification technique. The purpose of this approach is to extrapolate from our sample data how the entire population will vote in the Canadian general election using census data (Caetano, 2020c). For this analysis, we will use the 2017 General Social Survey as your census data. The goal is to calculate $\hat{y}^{PS}$, the estimate of our outcome variable after post-stratification using the following equation.

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

We can split the post-stratification tenique into three steps. First, we divide our sample data into j cells based on our predictor variables (Caetano, 2020c). We describe each bin by some combination of responses to our explanatory variables. This paper has three such variables: age, gender, and education. Therefore, one cell will contain all eighteen-year-old males with a bachelor degree. Another will hold all fifty-year-old women with a higher degree than the bachelor. These examples could continue as we enumerate through every possible combination of answers to our predictor variables.

Second, we estimate our response variable for each cell, $\hat{y}_j$ using our logistic regression model (Caetano, 2020c). For example, in Model 1, we estimate the probability that an individual will vote for the Liberals over other five parites. Third, we multiply the estimate by the population size, N_j, and divide by the total population size, $\sum N_j$ (Caetano, 2020c).

Results:

Table 3: Model 1 (Liberals' popular vote):

```
## # A tibble: 1 x 1
##   Liberal_popular_vote
##                  <dbl>
## 1                0.313
```

Table 4: Model 2 (Conservatives' popular vote):

```
## # A tibble: 1 x 1
##   Conservative_popular_vote
##                       <dbl>
## 1                     0.382
```

Table 5: Model 3 (NDP's popular vote):

```
## # A tibble: 1 x 1
##   NDP_popular_vote
```

```
##              <dbl>
## 1            0.146
```

Table 6: Model 4 (Bloc_Québécois' popular vote):

```
## # A tibble: 1 x 1
##   Bloc_Québécois_popular_vote
##                         <dbl>
## 1                      0.0402
```

Table 7: Model 5 (Green's popular vote):

```
## # A tibble: 1 x 1
##   Green_popular_vote
##                <dbl>
## 1             0.0998
```

Table 8: Model 6 (People's Party popular vote):

```
## # A tibble: 1 x 1
##   People_popular_vote
##                 <dbl>
## 1    0.00000000000290
```

Table 9: Summary of Post-Stratified Model Results

Model	Outcome Variable	Predicted Outcome (Popular Vote)	Actual Outcome (Popular Vote)
Model 1	Liberals	(31.318%)	(33%)
Model 2	Conservatives	(38.194%)	(34%)
Model 3	NDP	(14.591%)	(16%)
Model 4	Bloc Québécois	(4.025%)	(8%)
Model 5	Green Party	(9.985%)	(7%)
Model 6	People's Party	(0%)	(2%)

Overall, our model 1 predicted that the Liberal Party of Canada would receive 31.318% of the popular vote if every eligible voter votes on election day. The model 2 predicted that the Conservative Party would receive 38.194% if the popular vote if every eligible voter votes in the 2019 election. Likewise, the NDP would receive 14.591%; Bloc Québécois would receive 4.025%; the Green Party would receive 9.985% while we can almost ignore the popular vote that the People's Party would recive in percentage as it is less than 0.001%.

Our analysis shows that the Conservatives would win in a more clear margin against the Liberals in the popular vote if every eligible voter voted in the 2019 General Election, as the

margin in the analysis increases to 6.876% compared with 1% in the real election. This suggests us that our assumption and methodology have some weaknesses and needs to improve in the future data analysis.

Discussion:

In this analysis, we choose 2017 General Social Survey (GSS) and 2019 Canadian Election Study (CES) as census data and survey data, respectively. To count every eligible vote, we exclude respondents below 18 and respondents who refused to answer their preferred party or they would not vote on election day at all. We consider the remaining respondents are "good" voters in this analysis. Thus, we extract three same variables from census and survey data - age, gender, and education to create a multilevel logistic regression to predict if the Liberals Party could win the popular vote. The result is opposite to our assumption, as the Conservatives would win a bigger margin by 6.88% over the Liberals.

One area that surprises us is that in model 1 the predictor whose education is higher than the bachelor level seems not to have a strong correlation to the voting result in our analysis. The p-values of this case is 0.329. This may suggest that people who receive higher eduction have their own perspectives to the Liberal Party.

Overal, our six models shows that the Conservative will have a clear win in popular vote given the fact that every eligible voter voted and people who refused to vote and are undecided excluded. However, in Canadian electoral system, winning popular vote sometimes does not mean the party will win in the election, as the country's rule indicates that the political party who wins most electoral seats in the House of Commons will be able to form a government. Thus, to further predict if the Conservatives will have the right to the form a government, we will need to estimate the number of constituencies that the Conservatives could win. This further strategy will be discussed in the next step.

Weaknesses

This analysis shows some weaknesses during variable selection and data manipulation. First of all, we select 2017 General Social Survey as our census data which is an outdate dataset conducted by Statistics Canada, because the outdated information in this dataset may affect the accuracy of our prediction. However, this dataset is the latest one we can get as Statistics Canada implements General Social Survey every five year, so the next comprehensive survey will be in 2022. In the future, we could use other datasets as census data which are conducted in a more recent period.

Another limitation we can think of is dealing with missingless. In the analysis, we chose to remove cases who answered not to vote or were undecided. This strategy could allows either political party to lose some votes because undecided voters may lean to a certain political party during voting. Thus, there will be various of methods that we could potentially decrease the effect of missingness, such as assigning undecided based on the existing result.

In addition, predicting the winner in popular vote does not guaratee this party woould have the right to form a government. The reality can be in 2019 the Liberal lost the popular vote to the Conservatives but formed a minority government. We could estimate the number of electoral seats a party could achieve in the election given we have more detailed data in each constituency.

Lastly, using three predictors is not adequete to predict the popular vote a party could achieve. Instead, we can include more variables, such as income and reading preference to have better understanding in voters' voting habits so that our prediction can be more accurate.

Next Steps

Table 10: Distribution of Canadian Federal constituency

Province/Territory	Number of constituency
Alberta	34
British Columbia	42
Manitoba	14
New Brunswick	10
Newfoundland and Labrador	7
Nova Scotia	11
Ontario	121
Prince Edward Island	4
Quebec	78
Saskatchewan	14
Northwest Territories	1
Yukon	1
Nunavut	1
Total	338

Table 10 illustrates the distribution of the electoral seats in 10 provinces and 3 territories in Canada. To win a majority, a single political party needs 170 seats.

As mentioned in the weaknesses, we could further estimate the number of electoral seats that a political party could achieve in the 2019 general election. This will tell us if Liberals or Conservatives or other party in Canda could form a type of government, such as majority, minority or coalition government.

To do this, we will include the predictor 'province' as mentioend earlier, because this predictor would allow us to know the voter's residential palce. Later, we will group all eligible voters in the same province to impletement a multilevel regression, with post-stratification technique in each province and territory. The result would tell us the percentage of popular vote that a party could win in a province and we assign their

electoral seats based off their provincial popular vote. If a party wins at least 170 seats, it can form a majority government. If a party wins most electoral seats but not a majority, the party could form a minority government or unite other parties to form a coalition government.

In addition, a larger sample size is required in both census and survey datasets, because a larger sample size allows us to predict population more accurately.

Reference:

Caetano, S. (2020a, October 5). *Introduction to Logistic Regression* [PowerPoint Slides]. Quercus. https://q.utoronto.ca/courses/184060/files/9309406?module_item_id=1855302

Caetano, S. (2020b, October 12). *Introduction to Bayesian Inference* [PowerPoint Slides]. Quercus. https://q.utoronto.ca/courses/184060/files/9490196?module_item_id=1872494

Caetano, S. (2020c, October 19). *Multilevel Regression and Post-Stratification* [PowerPoint Slides]. Quercus. https://q.utoronto.ca/courses/184060/files/9490196?module_item_id=1872494

Canadian Election Study (2019). Welcome to the 2019 Canadian Election Study. http: //www.ces-eec.ca/.

"Election 2019 Results". 2019. The Globe And Mail. https://www.theglobeandmail.com/politics/election2019/results/.

Grenier, Éric. 2019. "How Voter Turnout Might Affect Who Wins In October". CBC. https://www.cbc.ca/news/politics/grenier-youth-turnout-1.5256600.

Hayes, M. (2019, October 22). Retrieved from Federal election 2019: Voter turnout dips to 66 per cent compared with enthusiasm that brought Trudeau to power four years ago: https://www.theglobeandmail.com/politics/article-federal-election-2019-voter-turnout-dips compared-with-enthusiasm/

Statistics Canada. 2019. "Surveys And Statistical Programs - General Social Survey - Family (GSS)". Www23.Statcan.Gc.Ca. https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&Id=335816.

Stephenson, Laura B; Harell, Allison; Rubenson, Daniel; Loewen, Peter John, 2020, "2019 Canadian Election Study - Online Survey Technical Report and Codebook.pdf", 2019 Canadian Election Study - Online Survey, https://doi.org/10.7910/DVN/DUS88V/HRZ21G, Harvard Dataverse, V1