

Re-evaluating 2019 Canadian Federal Election

Wei-chieh Li 1004548741

Dec 16, 2020

##Abstract

In this project, I will try to predict the 2019 Canadian Federal Election result if every citizen has voted. I will use the MRP model to complete this project. By separating people into groups, I will use the census data to calculate the total number of votes for each party, then predict the winner.

##Keywords

Linear Regression Model, Post-Stratification, Election, Canadian Election

##Introduction

The 2019 Canadian Federal Election is an election for all Canadian citizens to choose the government for the next few years. The Liberal party, led by incumbent Prime Minister Justin Trudeau, has won the election once again. Every Canadian citizen over the age of 18 has the right to vote in this election. The Liberal party, led by Justin Trudeau, won the majority seats, but lost the majority vote.

People may wonder how the results would change if every citizen with the right to vote has voted. Not every citizen has voted in the 2019 Canadian Federal election, we use the survey data to analyze how many more votes could've been received.

In this project, I will use the post-stratification technique to separate people into different cell groups. These cell groups are mainly based on how the education level and the age of respondents.

##Methodology

#Data

The survey data comes from the Harvard database on 2019 Canadian Election Study. There are two surveys, and I focus on the online survey in this project. Out of the various variables available, I chose the age, education level, and province of the respondent as the independent variable. On the other hand, the vote choice is the variable to predict. The age is the age of the respondent. The education level is the highest diploma that the respondent received. For example, if the respondent went to college but didn't complete, they will be categorized as high school education level. The province is the location that the respondent is currently living.

The vote choice is somehow confusing. Since there are multiple smaller parties in Canada, it is unlikely for them to win the election. For analyzing the final winner, we

eliminate the respondents that chose to vote to parties other than the Conservatives or the Liberals. In table 1, number 1 means that the respondent votes for the Liberals, while 0 represents the respondent votes for the Conservatives.

The census data comes from the Canadian 2016 census survey. It can be found on the official Canadian government website. By labelling them by their education levels, the respondents are grouped by their living province and education level. The 'counts' column represents that the total number of population in that province with the same education level. The estimate is the likeliness that they will either vote for the Liberals or the Conservatives. If the estimate number is lower than 0.5, the group will likely to vote for the Conservatives, while greater than 0.5 means that they will likely to vote for the Liberals. The Age label here is quite tricky. Since the 2016 Census data only gives a certain range of age, putting people into a category with plus or minus five years old. Here we take a mean of the range, since we cannot know the exact median or mean of every respondent. For example, the label 29.5 means the respondent can be in the age between 25 to 34.

#Table 1

Age	education_level	Province	cps19_votechoice	19	High School	Ontario	1
21	High School	Ontario	0				
20	High School	Ontario	1				
21	High School	Ontario	1				
19	High School	Ontario	1				
21	High School	Ontario	0				

#Table 2

Age	Province	education_level	count	estimate	prediction	29.5	Newfoundland and Labrador
Below high school	4585	0.4601567	Conservative	29.5	Newfoundland and Labrador		
High School	12790	0.4936242	Conservative	29.5	Newfoundland and Labrador		
College	7135	0.4743293	Conservative	29.5	Newfoundland and Labrador		
College	15935	0.4743293	Conservative	29.5	Newfoundland and Labrador		
College	1265	0.4743293	Conservative	29.5	Newfoundland and Labrador		
At or above Bachelor	15715	0.6051205	Liberal				

#Model

By the census data and the survey data above, we can create a linear regression model. The formula is given by:

$$y = \beta_0 + \beta_1 x_{age} + \beta_2 + \beta_3 + \beta_4 + \epsilon$$

Here, y represents the probability that the respondent will vote for either respondent. If y is greater than 0.5, we assume that the respondent will vote for the Liberals. β_0 represents the baseline. Since we have a categorical data, the baseline is set to "At or above Bachelor". β_1 here is the age of the respondent. β_3, β_4 , and β_5 represents if the respondent has no diploma, college diploma, and high school diploma respectively.

To perform analysis on the model, we put in the census data and analyze how many people will vote for each party. I use the post-stratification technique in this step. First, we predict the voting intentions by the survey data. I used the survey data to get the voting intentions of groups of people. Then, I categorized the population into groups and calculate the amount of population with the same education level and age. Next, I put the census data on population and predict their voting intentions. Finally, by calculating the total votes for each party, we predict the final winning party.

#Results

By the linear regression model, we can get an estimate for the census data.

Based on the linear regression model and the post-stratification data, the intercept gives a value of 0.640, which is closer to 1. The intercept shows that a person with a higher education level at the age of 0 will vote for. As the age of the respondent becomes higher, they are less likely to vote for the Liberals. When the respondent has a lower education level, they are less likely to vote for the Liberals as well.

If we look at the p-value, we can see that each is significant. Every educational level can significantly change value of y. There are little difference between education level lower than the Bachelor degree.

Call: `lm(formula = cps19_votechoice ~ Age + education_level, data = survey_data)`

Residuals: Min 1Q Median 3Q Max -0.6174 -0.4708 0.3884 0.5058 0.6213

Coefficients: Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.6397098 0.0128045 49.960 < 2e-16 **Age -0.0011725 0.0002284 -5.134 2.86e-07** education_levelBelow high school -0.1449638 0.0188114 -7.706 1.37e-14 **education_levelCollege -0.1307912 0.0101582 -12.875 < 2e-16** education_levelHigh School -0.1114962 0.0086536 -12.884 < 2e-16 *** — Signif. codes: 0 '0.001' '0.01' '0.05' '.' 0.1 ' ' 1

Residual standard error: 0.496 on 17629 degrees of freedom Multiple R-squared: 0.01615, Adjusted R-squared: 0.01593 F-statistic: 72.34 on 4 and 17629 DF, p-value: < 2.2e-16

[1] "Conservatives Win"

We can see that by the linear regression model and post-stratification technique, the conservatives win the majority vote by a great number.

##Discussion

#Summary

From the beginning, I categorized people into groups by their age and educational level. I eliminated the respondents with voting intentions party other than the

Liberals and the Conservatives. Also, I eliminated the respondents that refuse to answer their education level. Then I used the census data to predict the voting intentions of every Canadian Citizen by putting them into a certain cell. Finally, I predict the total vote that each party will receive, and also predict the final winner.

#Conclusion

The linear regression model shows a significant drop in support rate of the Liberal Party on people with lower educational level. This may be due to the political views of the Liberal Party. The Conservatives have the support of the less educated groups, as they support the more conservative ideas. The older population is more likely to have the same conservative idea as well. The census data doesn't include the population at the age 18-24 years old, which is the main supporters of the Liberal party. Only the highly educated population will vote for the Liberals, and they are still the minority in this project.

The final results is slightly unexpected, but it also follows the pattern on the actual election. The conservatives still won the majority votes. The part that is unexpected is the seats. By winning more than two times the vote, the Conservatives surely will win the majority seats as well. If the conservatives do win the election by this number, the whole current government has to change. By winning the majority seats, the Conservatives will have a dominating voice in the Parliament, making the Liberals very passive in decision making.

#Weakness & Next Step

As mentioned above, the census data somehow limits this project. There are not enough young people in the census data for us to analyze. If we include the respondents of the age 15-24 in the census data, it would also be inaccurate, since there are population at the age that is not enough to vote. Another weakness is that I eliminated many respondents in the first step. The Liberals formed an alliance with another party after the actual election, but that is not an option in this project, since there are no third choice. Additionally, more respondents were eliminated when they didn't answer their actual education level.

Another factor that may cause error is that the census data is recorded on 2016. We have to assume that there were no population change for the three year gap between the census survey and the CES election survey. Finally, the census data doesn't give an actual age for the respondents. This may cause an error on the analysis, since age is one of our deciding factor. If we can obtain the actual age, it is more likely to get a more accurate result. These two errors can be solved when the 2021 census data comes out. If we can have an up-to-date census data, it will definitely make this project better.

#Acknowledge 2019 Canadian Election study by Harvard Dataverse.
{DVN/DUS88V_2020, author = {Stephenson, Laura B and Harell, Allison and
Rubenson, Daniel and Loewen, Peter John}, publisher = {Harvard Dataverse}, title =
{{2019 Canadian Election Study - Online Survey}}, year = {2020}, version = {DRAFT
VERSION}, doi = {10.7910/DVN/DUS88V}, url =
{<https://doi.org/10.7910/DVN/DUS88V>}}