

# Study of 2019 Canadian federal election

Jiarong Ye

12/18/2020

## Abstract

The 2019 Canadian federal election is predicted in this report. However, Immigration is a major part of the Canadian population, they played important role in Canada's economy, but they are not allow to vote unless they obtain Canadian citizenship. In this report it will deeply analyze how permanents would affect the final result of 2019 Canadian federal election if they have to right to vote. The terminologies that would be used in this report in order to help us to analyze the data are logistic regression model and Post-Stratification.

Keywords: Logistic regression model, Post-Stratification, Permanent residents

## Introduction

The 2019 Canadian federal election was held on October 21, 2019. In result, Liberal party has popular vote 6,018,728 and 33.12% of total vote. Conservative party however has popular vote 6,239,227 and 34.34% of the total vote. It is obvious that this time Liberal party lost to Conservative party, which marks only the second time in Canadian history that a governing party formed a government while receiving less than 35 per cent of the national popular vote. The Liberals received the lowest percentage of the national popular vote of a governing party in Canadian history.

However, not everyone from Canada had voted. Some people might forgot to vote, or they chose to not vote. There is another group of people, they live and work in Canada just like any other Canadians. They are permanent residents but they are not allowed to vote. In this report, the issue, would the 2019 Canadian federal election be different if everyone in the Canada had voted including all the permanent residents, will be investigated.

In this report, I will be also analyzing how people's education, age and province would affect their vote choice. Since people in different education and age might have different thoughts on politics. People from different province have different preference in parties. This is based on what benefits the party have brought or will bring to the province.

I will use logistic regression model and Post-Stratification to show my prediction of the vote result. Logistic regression model is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables. Post-Stratification then Post-stratification is a common technique in survey analysis for incorporating population distributions of variables into survey estimates.

## Data

The survey data used in this project is from 2019 Canadian Election Study website. It is a reliable source because throughout its long history, the CES has been a rich source of data on Canadian's political behavior

and attitude. The leader of 2019 Canadian Election Study are Laura Stephenson, Allison Harell, Daniel Rubenson and Peter Loewen. They are all really professional at the election study. The data is online survey based data and has total sample size of 37,822. The online survey was conducted to document the attitudes of Canadians during and after the 2019 election. The target population includes all ages (18-year-old above) of Canadians and permanent residents across all provinces in Canada.

The census data is the highest level of educational attainment (general) by sex and selected age groups in 2016. It is from Statistics Canada, it is the national statistical office that ensures Canadians have the key information on Canada's economy, society and environment that they require to function effectively as citizens and decision makers. The targeted population are all 15 years old and over people for both sexes across Canada.

In order to perform Post-Stratification, it is necessary to keep all the variable names from both survey data and census data the same. The topic of the project is focused on both Canadians citizens and permanent residents, thus in survey data vote choice for both groups of people are combined together. The survey data only shows which year the person was born but the census data shows the age range. In order to keep them same, the year from survey data is converted to the age range it belonged to. However, the minimum age range included in census data is from 25 to 34. Therefore, the survey data could not include people from 18 years old to 24 years old then cases with age from 18 years old to 24 years old were eliminated.

## Model

### Model Specifics

I will be using two logistic regression models to model the proportion of voters. One logistic regression model will show who will vote for Liberal Party and another logistic regression model will show who will vote for Conservative party. The reason why the logistic regression model is used because the Y variable is binary meaning either vote for the one of the Liberal or conservative party or not. In one model, value 1 represents vote for Liberal party and value 0 represents not vote for Liberal party. Then in another model, value 1 represents vote for Conservative party and value 0 represents not vote for Conservative party. I will be using age range, gender, province and their education level to model the probability of voting for Liberal or Conservative party.

The logistic regression model I am using for both parties is:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_{Age35-44} + \beta_2 x_{Age45-54} + \dots + \beta_4 x_{Female} + \beta_5 x_{BC} + \beta_6 x_{MB} + \dots + \beta_{17} x_{highschool} + \beta_{18} x_{college} + \dots + \epsilon$$

p represents the probability of the event of interest occurring, here it refers to the probability that everyone in Canada including permanent residents will vote for Liberal or conservative party in 2019. Similarly,  $\beta_0$  represents the intercept of the model, and is the probability of voting for Liberal or conservative party for male people in age 25-34 from Alberta who do not own any diploma.

Model1: the logistic regression model for probability of people will vote for liberal party:

```
##
## Call:
## glm(formula = vote_liberal ~ Age + Sex + Province + education,
##      family = "binomial", data = reduced_survey)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3749  -0.9514  -0.7725   1.2786   2.1669
##
## Coefficients:
```

##	Estimate
## (Intercept)	-1.73635
## Age35 to 44	0.09270
## Age45 to 54	0.05071
## Age55 to 64	0.21862
## SexFemale	0.12311
## ProvinceBritish Columbia	0.74972
## ProvinceManitoba	0.72277
## ProvinceNew Brunswick	1.06740
## ProvinceNewfoundland and Labrador	1.54227
## ProvinceNorthwest Territories	1.18135
## ProvinceNova Scotia	1.38757
## ProvinceNunavut	1.01681
## ProvinceOntario	1.08213
## ProvincePrince Edward Island	1.21088
## ProvinceQuebec	0.97757
## ProvinceSaskatchewan	-0.21730
## ProvinceYukon	0.45772
## educationSecondary (high) school diploma or equivalency certificate	-0.34430
## educationCollege, CEGEP or other non-university certificate or diploma	-0.16265
## educationUniversity certificate or diploma below bachelor level	0.13437
## educationUniversity certificate, diploma or degree at bachelor level or above	0.30544
##	Std. Error
## (Intercept)	0.14806
## Age35 to 44	0.04421
## Age45 to 54	0.04530
## Age55 to 64	0.03786
## SexFemale	0.02754
## ProvinceBritish Columbia	0.06172
## ProvinceManitoba	0.07966
## ProvinceNew Brunswick	0.09919
## ProvinceNewfoundland and Labrador	0.11009
## ProvinceNorthwest Territories	0.49846
## ProvinceNova Scotia	0.09109
## ProvinceNunavut	0.52151
## ProvinceOntario	0.05151
## ProvincePrince Edward Island	0.21135
## ProvinceQuebec	0.05506
## ProvinceSaskatchewan	0.10368
## ProvinceYukon	0.44671
## educationSecondary (high) school diploma or equivalency certificate	0.14091
## educationCollege, CEGEP or other non-university certificate or diploma	0.13891
## educationUniversity certificate or diploma below bachelor level	0.14359
## educationUniversity certificate, diploma or degree at bachelor level or above	0.13828
##	z value
## (Intercept)	-11.727
## Age35 to 44	2.097
## Age45 to 54	1.119
## Age55 to 64	5.774
## SexFemale	4.470
## ProvinceBritish Columbia	12.148
## ProvinceManitoba	9.073
## ProvinceNew Brunswick	10.762
## ProvinceNewfoundland and Labrador	14.009

```

## ProvinceNorthwest Territories                2.370
## ProvinceNova Scotia                        15.232
## ProvinceNunavut                          1.950
## ProvinceOntario                          21.010
## ProvincePrince Edward Island              5.729
## ProvinceQuebec                           17.755
## ProvinceSaskatchewan                     -2.096
## ProvinceYukon                            1.025
## educationSecondary (high) school diploma or equivalency certificate -2.443
## educationCollege, CEGEP or other non-university certificate or diploma -1.171
## educationUniversity certificate or diploma below bachelor level      0.936
## educationUniversity certificate, diploma or degree at bachelor level or above 2.209
## Pr(>|z|)
## (Intercept)                               < 2e-16
## Age35 to 44                               0.0360
## Age45 to 54                               0.2630
## Age55 to 64                               7.73e-09
## SexFemale                                 7.82e-06
## ProvinceBritish Columbia                 < 2e-16
## ProvinceManitoba                        < 2e-16
## ProvinceNew Brunswick                   < 2e-16
## ProvinceNewfoundland and Labrador       < 2e-16
## ProvinceNorthwest Territories            0.0178
## ProvinceNova Scotia                     < 2e-16
## ProvinceNunavut                        0.0512
## ProvinceOntario                         < 2e-16
## ProvincePrince Edward Island            1.01e-08
## ProvinceQuebec                          < 2e-16
## ProvinceSaskatchewan                    0.0361
## ProvinceYukon                          0.3055
## educationSecondary (high) school diploma or equivalency certificate 0.0146
## educationCollege, CEGEP or other non-university certificate or diploma 0.2416
## educationUniversity certificate or diploma below bachelor level      0.3494
## educationUniversity certificate, diploma or degree at bachelor level or above 0.0272
##
## (Intercept)                               ***
## Age35 to 44                               *
## Age45 to 54                               ***
## Age55 to 64                               ***
## SexFemale                                 ***
## ProvinceBritish Columbia                 ***
## ProvinceManitoba                        ***
## ProvinceNew Brunswick                   ***
## ProvinceNewfoundland and Labrador       ***
## ProvinceNorthwest Territories            *
## ProvinceNova Scotia                     ***
## ProvinceNunavut                          .
## ProvinceOntario                         ***
## ProvincePrince Edward Island            ***
## ProvinceQuebec                          ***
## ProvinceSaskatchewan                     *
## ProvinceYukon
## educationSecondary (high) school diploma or equivalency certificate *
## educationCollege, CEGEP or other non-university certificate or diploma

```

```

## educationUniversity certificate or diploma below bachelor level
## educationUniversity certificate, diploma or degree at bachelor level or above *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 32942  on 25552  degrees of freedom
## Residual deviance: 31691  on 25532  degrees of freedom
## AIC: 31733
##
## Number of Fisher Scoring iterations: 4

```

Model2: logistic regression model for probability of people will vote for conservative party:

```

##
## Call:
## glm(formula = vote_conservative ~ Age + Sex + Province + education,
##      family = "binomial", data = reduced_survey)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6912  -0.8925  -0.6686   1.1060   2.1818
##
## Coefficients:
##
##                                     Estimate
## (Intercept)                        0.271051
## Age35 to 44                        0.286428
## Age45 to 54                        0.446109
## Age55 to 64                        0.473071
## SexFemale                         -0.399140
## ProvinceBritish Columbia          -1.383823
## ProvinceManitoba                  -0.897889
## ProvinceNew Brunswick             -1.513103
## ProvinceNewfoundland and Labrador -1.740355
## ProvinceNorthwest Territories     -2.134189
## ProvinceNova Scotia               -1.860545
## ProvinceNunavut                   -0.787109
## ProvinceOntario                   -1.339363
## ProvincePrince Edward Island      -1.955375
## ProvinceQuebec                    -2.128405
## ProvinceSaskatchewan              -0.351587
## ProvinceYukon                     -1.473773
## educationSecondary (high) school diploma or equivalency certificate  0.412428
## educationCollege, CEGEP or other non-university certificate or diploma 0.351063
## educationUniversity certificate or diploma below bachelor level      -0.026421
## educationUniversity certificate, diploma or degree at bachelor level or above 0.001058
##
##                                     Std. Error
## (Intercept)                        0.156350
## Age35 to 44                        0.047380
## Age45 to 54                        0.047723
## Age55 to 64                        0.040509
## SexFemale                          0.028480
## ProvinceBritish Columbia          0.055257

```

## ProvinceManitoba	0.071456
## ProvinceNew Brunswick	0.101467
## ProvinceNewfoundland and Labrador	0.122132
## ProvinceNorthwest Territories	0.641501
## ProvinceNova Scotia	0.101171
## ProvinceNunavut	0.509426
## ProvinceOntario	0.043794
## ProvincePrince Edward Island	0.250997
## ProvinceQuebec	0.051954
## ProvinceSaskatchewan	0.076669
## ProvinceYukon	0.447765
## educationSecondary (high) school diploma or equivalency certificate	0.151854
## educationCollege, CEGEP or other non-university certificate or diploma	0.150343
## educationUniversity certificate or diploma below bachelor level	0.156022
## educationUniversity certificate, diploma or degree at bachelor level or above	0.150138
##	z value
## (Intercept)	1.734
## Age35 to 44	6.045
## Age45 to 54	9.348
## Age55 to 64	11.678
## SexFemale	-14.015
## ProvinceBritish Columbia	-25.043
## ProvinceManitoba	-12.566
## ProvinceNew Brunswick	-14.912
## ProvinceNewfoundland and Labrador	-14.250
## ProvinceNorthwest Territories	-3.327
## ProvinceNova Scotia	-18.390
## ProvinceNunavut	-1.545
## ProvinceOntario	-30.583
## ProvincePrince Edward Island	-7.790
## ProvinceQuebec	-40.967
## ProvinceSaskatchewan	-4.586
## ProvinceYukon	-3.291
## educationSecondary (high) school diploma or equivalency certificate	2.716
## educationCollege, CEGEP or other non-university certificate or diploma	2.335
## educationUniversity certificate or diploma below bachelor level	-0.169
## educationUniversity certificate, diploma or degree at bachelor level or above	0.007
##	Pr(> z )
## (Intercept)	0.082986
## Age35 to 44	1.49e-09
## Age45 to 54	< 2e-16
## Age55 to 64	< 2e-16
## SexFemale	< 2e-16
## ProvinceBritish Columbia	< 2e-16
## ProvinceManitoba	< 2e-16
## ProvinceNew Brunswick	< 2e-16
## ProvinceNewfoundland and Labrador	< 2e-16
## ProvinceNorthwest Territories	0.000878
## ProvinceNova Scotia	< 2e-16
## ProvinceNunavut	0.122325
## ProvinceOntario	< 2e-16
## ProvincePrince Edward Island	6.68e-15
## ProvinceQuebec	< 2e-16
## ProvinceSaskatchewan	4.52e-06

```

## ProvinceYukon 0.000997
## educationSecondary (high) school diploma or equivalency certificate 0.006609
## educationCollege, CEGEP or other non-university certificate or diploma 0.019539
## educationUniversity certificate or diploma below bachelor level 0.865528
## educationUniversity certificate, diploma or degree at bachelor level or above 0.994376
##
## (Intercept) .
## Age35 to 44 ***
## Age45 to 54 ***
## Age55 to 64 ***
## SexFemale ***
## ProvinceBritish Columbia ***
## ProvinceManitoba ***
## ProvinceNew Brunswick ***
## ProvinceNewfoundland and Labrador ***
## ProvinceNorthwest Territories ***
## ProvinceNova Scotia ***
## ProvinceNunavut
## ProvinceOntario ***
## ProvincePrince Edward Island ***
## ProvinceQuebec ***
## ProvinceSaskatchewan ***
## ProvinceYukon ***
## educationSecondary (high) school diploma or equivalency certificate **
## educationCollege, CEGEP or other non-university certificate or diploma *
## educationUniversity certificate or diploma below bachelor level
## educationUniversity certificate, diploma or degree at bachelor level or above
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 32629  on 25552  degrees of freedom
## Residual deviance: 29921  on 25532  degrees of freedom
## AIC: 29963
##
## Number of Fisher Scoring iterations: 4

```

## Post-Stratification

The final prediction of probability of everyone in Canada (including permanent residents) voting for Liberal party by Post-Stratification method:

```

## Rows: 520
## Columns: 6
## $ Age      <chr> "25 to 34", "25 to 34", "25 to 34", "25 to 34", "25 to 34..."
## $ Sex      <chr> "Male", "Male", "Male", "Male", "Male", "Female", "Female..."
## $ Province <chr> "Newfoundland and Labrador", "Newfoundland and Labrador",...
## $ education <chr> "No certificate, diploma or degree", "Secondary (high) sc..."
## $ count    <dbl> 2615, 6440, 7030, 645, 5800, 1970, 6345, 8910, 620, 9910,...
## $ estimate <dbl> 0.4516309, 0.3685634, 0.4117507, 0.4850756, 0.5278111, 0....

## # A tibble: 1 x 1

```

```
## alp_predict
##      <dbl>
## 1      0.328
```

The final prediction of probability of everyone in Canada (including permanent residents) voting for conservative party by Post-Stratification method:

```
## Rows: 520
## Columns: 6
## $ Age      <chr> "25 to 34", "25 to 34", "25 to 34", "25 to 34", "25 to 34..."
## $ Sex      <chr> "Male", "Male", "Male", "Male", "Male", "Female", "Female..."
## $ Province <chr> "Newfoundland and Labrador", "Newfoundland and Labrador",...
## $ education <chr> "No certificate, diploma or degree", "Secondary (high) sc..."
## $ count    <dbl> 2615, 6440, 7030, 645, 5800, 1970, 6345, 8910, 620, 9910,...
## $ estimate <dbl> 0.1870484, 0.2579069, 0.2463378, 0.1830640, 0.1872094, 0....

## # A tibble: 1 x 1
## alp_predict
##      <dbl>
## 1      0.332
```

## Results

Below is how Model1: the logistic regression model for probability of people will vote for liberal party could be expressed:

$$\log\left(\frac{p}{1-p}\right) = -1.7364 + 0.0927x_{Age35-44} + \dots + 0.1231x_{Female} + 0.7497x_{BC} + \dots - 0.3443x_{highschool} + \dots + \epsilon$$

$\hat{\beta}_0$  here is -1.1736. It is also a log of odds when all the x values are zero. All predictor variables here are categorical. The positive coefficient of  $x_{Age35-44}$  is 0.0927, it means that holding all other explanatory variables in the model fixed, person who is in age 35 to 44 has in average 0.0927 more possibility of voting for Liberal party than person who is in age 25 to 34. Comparing all the age ranges, ages from 55 to 64 has the biggest positive coefficient, it implies that 55 to 64 years old people have the most possibility to vote for Liberal party. The positive coefficient of  $x_{Female}$  is 0.1231, meaning holding all other explanatory variables in the model fixed, female person has in average 0.1231 more possibility of voting for Liberal party than male person. 0.7497 is a coefficient for  $x_{BC}$ . Person who is from British Columbia has in average 0.7497 more possibility of voting for Liberal party than person who from Alberta. Looking at other provinces, -0.2173 is the coefficient for Saskatchewan. It tells us that Person from Saskatchewan would have in average -0.2173 less possibility of voting for Liberal party than person who is from Alberta. Saskatchewan is also the only province with negative coefficient, thus people in Saskatchewan have the least voting rate to Liberal party among other provinces. Meanwhile, Newfoundland and Labrador has the biggest positive coefficient, 1.5423. It implies that Liberal party is most popular in Newfoundland and Labrador. Now let move on how education level could affect the possibility that person would vote for Liberal party. -0.3443 is the coefficient for person with high school diploma. Person whose highest education level is high school would have in average 0.3443 less possibility of voting for Liberal party than person who achieves no certificate or diploma. It is also the smallest coefficient comparing to other education level coefficients meaning people whose highest educational attainment is high school diploma would have the lowest possibility to choose Liberal party. However, people with university certificate have the biggest coefficient, 0.3494 among other educational attainments. Therefore, people with university certificate are most willing to choose for Liberal party. Now, logistic model is used to do Post-Stratification by taking advantage of demographics to extrapolate how entire population will vote. The table shows that it partition the census data into demographic cells. Each



cell is a demographic to a unique combination of gender, province, age and educational attainment. Each cell also includes the estimate response variable, the probability of Liberal party being chosen calculated by the logistic model introduced previously. How could we get the final prediction by Post-Stratification? We first use estimate of each cell times its corresponding population size and then take the sum of the value from every single one of the cell, lastly divide it by the entire population size. Here  $\hat{y}^{ps} = 0.3282$ . The prediction of probability of Liberal party being chosen in entire population size is 32.82%.

Now let's see Model2: logistic regression model for probability of people will vote for Conservative party. It could expressed as:

$$\log\left(\frac{p}{1-p}\right) = 0.2711 + 0.2864x_{Age35-44} + \dots - 0.3991x_{Female} - 1.3838x_{BC} + \dots + 0.4124x_{highschool} + \dots + \epsilon$$

In conclusion, 55 to 64 years old female people from Alberta whose highest educational attainment is high school diploma would have the greatest possibility of choosing Conservative party. However, 25 to 34 years old male people from Northwest Territories whose highest educational attainment is university diploma would have the lowest possibility of choosing Conservative party. After the Post-Stratification is done, the prediction of probability of Conservative party being chosen in entire population size is 33.14%.

## Discussion

I did data cleaning process for both survey and census data. I then created logistic regression model by using the survey data for both Liberal and Conservative parties. The two models is used to see the how could variables affect the probability of Liberal and Conservative parties being chosen. After, the Post-Stratification is started, the census data was divided into many demographic cell and each cell contains estimate response variable calculated based on the logistic regression model. Lastly, after some calculation the probability of Conservative parties being chosend in the entire population size is predicted.

By using Post-Stratification method, prediction of probability of Liberal and Conservative party being chosen is calculated, which are 32.82% and 33.14%. Compare to the official final percentage 33.12% and 34.34% published by Canadian government, the prediction percentage is very close to the final percentage. Overall, both Liberal and Conservative party percentage predictions are a little bit lower than the final percentage. If we disregard all the error made in the prediction, it could tell that most of the permanent residents would like to vote for other parties except for Liberal and Conservative party. The difference between the prediction and final percentages for both parties are: 33.12%-32.82%=0.3% (Liberal Party) and 34.34%-33.14%=1.2% (Conservative Party). The difference for Liberal party is relatively smaller than Conservative Party, thus permanent residents are more likely to choose Liberal than Conservative. This is not surprising because permanent residents in Canada are immigrants from all over the world. The leader of Liberal Party, Justin Trudeau has brought many immigrants and refugees to the country, Trudeau always wants Canada to be a multicultural Country and offers many benefits to the refugees and immigrants.

## Weakness

The biggest obstacle in the data cleaning process is that both data come from different years. The survey data is from 2019 but the census data is from 2016. The gap between 2019 and 2016 is 3 years and during these 3 years many things could be changed, like population, people's perspectives and so on. Therefore, the prediction generated by Post-Stratification would not that accurate. In this report we assume all permanent residents and Canadian citizens are allowed to vote. However, the census data we used includes Canadian citizens, landed immigrants (permanent residents) and non-permanent residents and their families living with them in Canada. It includes extra groups of people in the census people which would lead to inaccurate result by Post-Stratification. The census data only includes people from age 25 to 64. It does not include people from age 18 to 24 and 65 and above. The census data only included age range but not specific age, the age range is a categorical variable in this project. If the census data includes specific ages, the age would be a discrete predictor variable and would show specifically age in one year difference could affect the probability of choosing certain party.

## Preference

Clarke, Seán, and Cath Levett. “Canada Election 2019: Full Results.” The Guardian, Guardian News and Media, 23 Oct. 2019, [www.theguardian.com/world/2019/oct/22/canada-election-2019-full-results](http://www.theguardian.com/world/2019/oct/22/canada-election-2019-full-results).

Swaminathan, Saishruthi. “Logistic Regression - Detailed Overview.” Medium, Towards Data Science, 18 Jan. 2019, [towardsdatascience.com/logistic-regression-detailed-overview-46c4da4303bc](https://towardsdatascience.com/logistic-regression-detailed-overview-46c4da4303bc).

Little, R. J. A. Post-Stratification: A Modeler’s Perspective.