

Analysis of how everyone's voting behavior in 2019 Canadian Federal Election

Yuxin Li(1002330998)

21/12/2020

Code and data supporting this analysis is available at: <https://github.com/Vickyli6762/FinalAssessment>

Abstract

We estimate and analyze the results in 2019 Canadian Federal Election by MRP model based on the point view of Jack Beiley. And we selected two main party in Canadian Federal, Liberal party and Conservative Party. The final result shows the proportion of everyone's voting for each party respectively.

Keywords

MRP model, Post-stratification, Canadian Federal Election, Liberal Party, Conservative Party. CES, GGS.

Introduction

We are interested in voters' behavior of the 2019 Canadian Federal Election, thus we build a MRP model to analyze the vote outcome by using statistical analysis based on CES and a post-stratification dataset. Here we will analyze the two main Candidates Party, Liberal Party and Conservative Party. And the Post-stratification technique is applied to have multi-level modeling in this assignment.

Bailey noted that Liberal party won 296 tickets and Conservative party won 281 tickets in 2017. As the result shown the Liberal Party won the election in 2019, it is significant for next election that using statistic analysis investigate why Canadian citizens voted for Liberal Party. And by considering different gender groups, age levels, education levels, provinces and racial, we used "Liberal_estimate" and "Conservative_estimate" to represent each voter behavior during election.

Methodology

- Data

The models have two dataset, which is survey data and census data. And the survey data we used is 2019 Canadian Election Study (CES). Census data we selected is Canadian General Social Survey (GSS).

- Model Specifics

The models we used are logistic regression models to determine certain factors of voters that would have an effect on whether the two main Candidates, Liberal Party or Conservative Party would win the election, respectively. And five variables were taken into the consideration which are economic financial status, the education level, their age group, gender group and province. There are two models, one for the prediction of voting for Liberal Party, and the other for the prediction of voting for Conservative Party. Because logistic model can be used to model binary response variables, thus we chose logistic models which are suitable for the prediction of whether citizens vote for candidates. Two models were used, because they can reflect more comprehensive predictions of people's voting choice between Liberal Party and Conservative Party. The logistic models we used are:

$$\log \frac{p_L}{1 - p_L} = \beta_0 + \beta_1 X_{age} + \beta_2 X_{education} + \beta_3 X_{ecoself} + \beta_4 X_{province} + \beta_5 X_{gender}$$

$$\log \frac{p_C}{1 - p_C} = \beta_0 + \beta_1 X_{age} + \beta_2 X_{education} + \beta_3 X_{ecoself} + \beta_4 X_{province} + \beta_5 X_{gender}$$

- Where p_L is the probability of Liberal Party would win the election, and p_C is the probability of Conservative Party would win the election. β_0 represents the intercept of the model. $\beta_1, \beta_2, \beta_3, \beta_4$, and β_5 represent the change in log odds for every one unit increase in $X_{education}, X_{ecoself}, X_{age}, X_{province}$, and X_{gender} respectively. X_{age} is the voters' age group, which has four levels: 19-24, 25-44, 45-65, and older than 65. $X_{ecoself}$ represents whether the oneself financial situation made by the policies of Canadian Federal government better, worth or not much difference. $X_{province}$ represents voters living different province such as Ontario, British Columbia. The variable X_{gender} describes voters' gender, which is grouped into Female and Male. $X_{education}$ represents voter's education level, which includes completed secondary/high school, bachelor's degree, master's degree and some technical, community college, CEGEP, College.

```
##
## Call:
## glm(formula = voteforliberal ~ province + sex + age, family = "binomial",
##      data = survey_data)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.827  -1.286   0.788   1.012   1.440
##
## Coefficients:
##
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -0.7307495   0.0439664  -16.621 < 2e-16 ***
## provinceBritish Columbia    0.7395700   0.0436660   16.937 < 2e-16 ***
## provinceManitoba      0.5257176   0.0579285    9.075 < 2e-16 ***
## provinceNew Brunswick    0.7731000   0.0767514   10.073 < 2e-16 ***
## provinceNewfoundland and Labrador 0.6011186   0.0861953    6.974 3.08e-12 ***
## provinceNorthwest Territories 1.7500482   0.5013703    3.491 0.000482 ***
## provinceNova Scotia     0.9914225   0.0737255   13.447 < 2e-16 ***
## provinceNunavut        0.4528861   0.4027339    1.125 0.260789
## provinceOntario         0.7333045   0.0348708   21.029 < 2e-16 ***
## provincePrince Edward Island 1.0095155   0.1670882    6.042 1.52e-09 ***
## provinceQuebec         1.3627248   0.0394341   34.557 < 2e-16 ***
## provinceSaskatchewan    0.0610211   0.0633094    0.964 0.335118
## provinceYukon          1.0307081   0.3556028    2.898 0.003750 **
## sexMale                0.1068506   0.0222214    4.808 1.52e-06 ***
```

```

## age                                0.0072753  0.0006605  11.015  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 50472  on 37530  degrees of freedom
## Residual deviance: 48786  on 37516  degrees of freedom
## (291 observations deleted due to missingness)
## AIC: 48816
##
## Number of Fisher Scoring iterations: 4

##
## Call:
## glm(formula = voteforConservative ~ province + sex + age, family = "binomial",
##      data = survey_data)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.440  -1.012  -0.788   1.286   1.827
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      0.7307495  0.0439664  16.621  < 2e-16 ***
## provinceBritish Columbia -0.7395700  0.0436660 -16.937  < 2e-16 ***
## provinceManitoba      -0.5257176  0.0579285  -9.075  < 2e-16 ***
## provinceNew Brunswick -0.7731000  0.0767514 -10.073  < 2e-16 ***
## provinceNewfoundland and Labrador -0.6011186  0.0861953  -6.974 3.08e-12 ***
## provinceNorthwest Territories -1.7500482  0.5013703  -3.491 0.000482 ***
## provinceNova Scotia    -0.9914225  0.0737255 -13.447  < 2e-16 ***
## provinceNunavut        -0.4528861  0.4027339  -1.125 0.260789
## provinceOntario        -0.7333045  0.0348708 -21.029  < 2e-16 ***
## provincePrince Edward Island -1.0095155  0.1670882  -6.042 1.52e-09 ***
## provinceQuebec         -1.3627248  0.0394341 -34.557  < 2e-16 ***
## provinceSaskatchewan   -0.0610211  0.0633094  -0.964 0.335118
## provinceYukon          -1.0307081  0.3556028  -2.898 0.003750 **
## sexMale                -0.1068506  0.0222214  -4.808 1.52e-06 ***
## age                   -0.0072753  0.0006605 -11.015  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 50472  on 37530  degrees of freedom
## Residual deviance: 48786  on 37516  degrees of freedom
## (291 observations deleted due to missingness)
## AIC: 48816
##
## Number of Fisher Scoring iterations: 4

```

Results

We estimate the proportion of voters in favor of average voting for Liberal Party to be 0.647. This is based of our post-stratification analysis of the proportion of voters in favor of Liberal Party modelled by a MRP model, which accounted for age, gender, ecoself, education, and province.

In addition, We estimate the proportion of voters in favor of average voting for Conservative Party to be 0.391. This is based of our post-stratification analysis of the proportion of voters in favor of Conservative Party modelled by a MRP model, which accounted for age, gender, ecoself, education, and province.

Discussion

- Summary

In summary, we all know the final results in 2019 Canadian Federal Election, which Liberal Party of Canada won the majority of seats during the election. However, it is still important for preparation of next election to investigate the voters' behaviors. Hence, we discussed why most voters chose Liberal Party comparing Conservative Party in this assignment. And Logistic regression models could help us forecast the choice of citizens more accurately. Since Canada as a multicultural countries involves different province, class and age groups, we introduce post-stratification to investigate five variables from CES, which education levels, their age period, gender group, living province, and ecoself status.

- Conclusions

In conclusion, our model estimate the proportion of voting for Liberal Party to be 0.647 and Conservative Party to be 0.391 based on five variables, which means different age, gender group, different education levels, ecoself status and racial group might influence everyone's voting behavior. For example, as Liberal Party respecting the culture of Quebec, people who lives in Quebec province largely might be willing to vote for Liberal Party. Hence, the policies that each Party promised for age group, gender, or particular province during election have significant impact on the results of election.

- Weakness

Although the model analyze the election based on five variables, there are still many variables adding to this model. For example, the employment status of people might have influence on result. People who does not have job prefer one Party which promises more job seats adding. In addition, different racial groups could take consideration on the result as well. Since Canada as multicultural countries, there are different racial group such as Chinese, Japanese, and Americans. Also, according to special voting system in Canada, it does not mean that winning more tickets will win the selection. As data shown in 2019, Liberal party won 5,911,588 in total and Conservative Party won 6,150,177 in total, but Liberal party still won the election. The reason for this is that the election is not based on total tickets of each Party. The total seats are 228 in the whole Canada. And the Party which won the majority seats in the House will win the election. Therefore, Liberal Party won the election by 157 seats comparing 121 seats of Conservative Party.

- Next Steps

From weakness we know many other variables could be added in this model, therefore, we could add more variables such as racial groups, employment status to make the results more accurately. In addition, the survey data and census data should be more precise such as removing missing values. And if it is possible, we could compare 2019 Canadian Federal Election with other year election such as 2015 Canadian Federal Election and 2011 Canadian Federal Election.

Reference

- 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
- 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. <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DUS88V>
- Dataset: General Social Survey, Cycle 31: 2017: Family
- GutHub: <https://github.com/Vickyli6762/FinalAssessment>
- CBC news <https://newsinteractives.cbc.ca/elections/federal/2019/results/>