The importance of turnout in 2019 Canadian Federal Election result

Ziqi Gao, 1003051092

20/12/2020

The importance of turnout in 2019 Canadian Federal Election result

Ziqi Gao 1003051092

December 22th, 2020

Abstract:

The low voter turnout is a significance issue in election events that is an obvious reason in why the sample opinion has a huge different with the poll opinion. The impact of polictial participation is the research question in this paper. This paper builds a multilevel regression with post-stratification model investigates how age, sex, province and education level have influenced the result of 2019 Canadian Federal Election, by using the 2019 CES Data. From this model, we would see that the distribution of voter who prefer the Liberal Party and which group would choose the Conservative Party. By our model analysis, the Liberal Party would has a slightly higher support rate to win this election. With these implications from the model, we can analysis the influence of the low turnout rate, and discuss the limitation of this model and the data that we have collect. These results are mostly aligned with our common sense, however, more investigation can be done from sociological aspects by utilizing this model. Code and data supporting this analysis is available at: https://github.com/Christine-Gao1092/STA304-Final-Project

Keywords: 2019 Canadian Federal Election, Liberal, Political participation, Low voter turnout, limitation

Introduction:

The 2019 Canadian Federal Election has the highest voter turnout rate during 25 years, which is 65.95% of Canadian who has eligibility to vote for this election [7]. The political participation has played an important role in the election, and the vote turnout rate has significant influence to election consequence. Although the voter turnout rate is increasing since 20 century, the percentage of turnout still does not cover the majority of Canadian political position and their willingness. Due to the low voter turnout, the results of public opinions may have an obvious different with the sample result, even the final result. From the low voter turnover issue and the different between public willingness and actual result, the research question in this paper is that the result changing when everyone has voted.

In this paper, we first obtain the survey data set from 2019 CES Data, which is a survey data that collected online to ask the participators about their choice in 2019 Canadian Federal Election. This data set contain 620 variables and collected 37822 samples to ask about their background information, satisfaction about the party, the choice of Federal Election etc.[1]. However, in this

report, we have focused on their age, gender, education level, the province that they are living and their political choice, because we believe these variables that are more closely linked to the result in final election choice from common sense.

In order to do so, we decided to build a multilevel logistic model and employ the post-stratification technique with the previously mentioned models. We obtained our data from the 2019 CES Data [2] + General Social Survey 2017[3]. The reason we chose this method is that logistic regression, like all regression models, is a predictive analysis. In addition, logistic regression is used to describe data and to explain the relationship between a binary dependent variable and one or more independent variables.[4] However, in our model analysis, we need to not only consider the individual influences but also the group factor, which leads us to the multilevel logistic model method.

In the meantime, the technique of post-stratification allows us to separate the data so that we can see the pattern. And because of the characteristics of the Canadian election system, this technique helps us to predict the result more precisely. In the following subsections, we will discuss the details of our model and the application of post-stratification techniques.

Data:

‘cps19_votechoice’	‘cps19_yob’	‘cps19_gender’	‘education’	‘cps19_province’
Liberal Party :7522	Min. :18.00	A man: 10922	Length:26090	Ontario:10041
Conservative Party:7514	1st Qu.:37.00	A women :14983	Class :character	Quebec:5915
Don’t know :3724	Median :52.00	Other: 185	Mode :character	Alberta :3171
ndp :3485	Mean :50.95			British Columbia :2953
Green Party:1949	3rd Qu.:64.00			Manitoba:1143
Bloc Qu9b9cois :1248	Max. :99.00			Saskatchewan: 942
(Other): 648				(Other): 1925

The frame population is all Canadian citizens who has eligible to vote in 2019 Canadian Federal Election though online survey and phone survey. The target sample size is 37822 and the actual respond sample size 26090. The response rate is 69%. Respondents are mostly found by online survey, and in the end of this survey, it also have a question for people do not participate the Election to explain the reason why they does not vote in 2019 Canadian Federal Election. The interviewer would also explain the importance of the interviewee and encourage citizens to participate.

Among the total 620 variables, we selected “cps19_votechoice” as our dependent (Y) variable, and use “cps19_yob”, “cps19_gender”, “education” and “cps19_province” as our independent (X) variables to investigate the relations between a person’s age, sex, education level and living province when they choose the Party in Federal Election. In the final data frame, all “sex”,

"education" and "province" are categorical variables, only the "yob" is the numercial variable. From the above table, it illustrates that the mean age is approximately 51 years old, which means odder Canadian citizens would more likely to participate the Federal Election survey. In addition, Ontario is the province who has the most participants among the seven provinces, and the number of women participants has almost one and half larger than the men group. Moreover, the support rate between Liberal Party and Conservative Party do not have a siginificant difference in this survey result.

Model:

As mentioned above, we construct a multilevel logistic regression model using the following equations:

$$log\frac{p}{1-p} = \alpha_j + \beta_1\widehat{age} + \beta_2\widehat{gender} + \beta_3\widehat{education} + \epsilon$$

Where p represents the proportion of voters who will vote for Liberal Party. βs represent the slope of our dependent variables. α_j, however, represents the formula of our level 2 variable, presented by the following equation. ϵ is the error term for this estimation.

$$\alpha_j = \gamma_0 + \eta_1 Province + \mu$$

In this equation, γ is the intercept. η , which has a similar function with beta, acts as the slope for our dependent variable (province) and μ is the error term of this estimation.

We use the vote intention variable as our response variable(represented by the left side of the first equation), and age, education, gender, and household income as our explanatory and level 1 (individual) variables (represented by the right of the first equation). Our level 2 variable is race, which is a group variable(represented by the second equation). The ultimate goal is for our model to be able to answer the question like ' how does the probability for Trump to win the election change every additional level a person is educated' or 'does factors of age, gender, or income have influences on the probability of voting Trump'. In addition, the model can be seen as the training model for the testing model in the post-stratification which we will discuss in the next section.

Post-stratification

Post-stratification allows us to adjust the weights so that the totals in each group are equal to the known population totals.[5] In other words, it increases the precision of our final prediction. According to the Canadian election system, the majority party required at minimum 170 seats in the House of Commons, if the winner Party does not achieve to the majority requirement, it need to gain and maintain the confidence of Parliament and it need a supprt by other party when it want to issue and execute a new policy[6]. The seats is selected by the Canadian citizens in the same election zone, the candidate who has the most 338 seats in House of Commons would become the winner of the election [7]. Due to each province has different number of electoral votes, we wish to use the post-stratification to adjust the weight between each province so that the totals in each group equal to the population. By doing so, we use individual estimations that we got from our regression models in the previous section, sorted them to the state that they

belong to and computed a post-stratification weight for each state based on the 2019 Canadian CES Data.

Result:

The random intercept is given by different province, and the support to Liberal Party and Conservative Party in follows:

'Vote choice'	'Alberta'	'British Columbia'	'Manitoba'	'New Brunswick'	'Newfoundland and Labrador'	'North west Territories'	'Nova Scotia'	'Nunavut'	'Ontario'	'Prince Edward Island'	'Quebec'	'Saskatchewan'	'Yukon'
Liberal Party	-1.28277337	0.05892908	-0.27717711	0.40272215	0.7749473 1	0.72078776	0.83563821	0.34860003	0.25591115	0.64202062	0.75319121	-1.2758135 3	0.04149777
Conservative Party	1.28277519	-0.05892708	0.27717903	-0.40272009	-0.7749448 8	-0.72078175	-0.83563594	-0.34859673	-0.25590919	-0.64201777	-0.75318934	1.2758151 2	-0.04149564

In this model, all three variables (age, gender and education level) are significant predictors, as their p-values are extremely small.The AIC and BIC are quite large for the model, with values equal to 19051.4 and 19104.7, respectively. This suggests that this model might not be very predictable.

Summary:

The goal of our project is to predict who will win the Canadian Federal Election in 2019 by a multilevel logistic regression model. First of all, we have built a new data set by separating the supporters between Liberal Party and Conservative Party with three variables from the origin survey data on October 21th 2019, which is age, gender and education_level. In our logistic model, it represents that the people more likely to vote for Liberal Party as their age increase (β_1 = -0.004449), in the female group(β_2= 0.348226), the higher level of education (β_3=0.417738). On the other hand, the people who vote for Conservative Party are more likely in a younger age group and within a male group or people have a relative lower level of education. Moreover, by our post-stratification model, we divided the origin census data set into 7 groups by provinces variable to identify the states distribution between Liberal Party and Conservative Party, and plug the new census data set into a logistic regression model to calculate the probability each candidate can win this election. According to our post-stratification model, the estimate of $\hat{p}$ in Conservative Party is slightly higher than Liberal Party only in Alberta, Manitoba and Saskatchewan provinces. Therefore we can predict that the Liberal Party would has a slightly higher support rate in the 2019 Canadian Federal Election, which is as same as the realistic result.

Conclusion:

According to our post-stratification model, the mean estimated proportion of voters in favor of voting for Liberal Party is 0.6197, which is slightly higher than the mean of estimated proportion for Conservative Party. Although the overall vote between two candidates parties are quite similar, the first and third quartiles of support in Liberal Party are slightly higher than Conservative Party, but the Conservative Party has a higher support in minimum and maximum estimated proportion of voters in their choice. Based on these estimated proportions, we predict that Liberal Party will win the 2019 Canadian Federal Election.

Weakness:

There are several weaknesses of this survey and analysis. First of all, the variables examined by the survey are mostly categorical variables. Using categorical variables in a linear regression model works, but might limit the result compared to using continuous variables. Secondly, there are 13 provinces has included in our survey data set, and Ontario has occupied almost half of the participants, and provinces is the important variable when we analysis the province distribution of voters. The large proportion of Ontario interviewees may lead to a bias in the survey data result, for example, they may prefer the party which can lead more benefit to Ontario province rather than the Canada. Lastly, the census data is collected in 2017 rather than 2019, which means the data would not reflect the last public opinion about the party, because the support of Liberal in 2019 Election is lower than 2015 Election by real result. The old data set may cause bias as well.

Next Steps:

First of all, we should do is to collect larger and up-to-date data sets to be our census data because the sample size of census data is smaller than the survey data. With the latest knowledge and data set, it will give better predictions in our model. Another survey could be collected after the election, and comparison between our model's prediction and the survey's results could be made. Also, we should include more variables in our model in order to be more realistic, such as the income level of voter, their occupation and whether they are immigrants, because each party has different governing idea that has different beneficial group. By adding more relative variables can provide the evidence to classify the group of voters and find the reasons about why the election has a low turnout rate.

Reference:

1.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 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.Government of Canada, Statistics Canada. Population Estimates on July 1st, by Age and Sex,
Government of Canada, Statistics Canada, 22 Dece. 2020,
www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1710000501 4.Sommet, Nicolas, and Davide Morselli. "Correction: Keep Calm and Learn Multilevel Logistic Modeling: A Simplified Three-Step Procedure Using Stata, R, Mplus, and SPSS." International Review of Social Psychology,

vol. 30, no. 1, 8 Sept. 2017, pp. 229–230., doi:10.5334/irsp.162. 5.Statistics Solutions. “What Is Logistic Regression?” Statistics Solutions, 9 Mar. 2020, www.statisticssolutions.com/what-is-logistic-regression/. 6.Elections Canada. “Elections Step by Step”, https://electionsanddemocracy.ca/canadas-elections/canadas-election-process/elections-step-step 7.Krystyne Manzer. “2019 Canadian Election Results”, 22 Oct. 2019, https://www.rbcgam.com/en/ca/article/canadian-federal-election-2019/detail 8. Gao,Gui,Qian,Zhang. “Prediction on 2020 American Election”, 2nd Nov. 2020,https://github.com/KaixiZhang99/STA304-PS2-Group164 9.Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686 10. Hadley Wickham and Evan Miller (2020). haven: Import and Export ‘SPSS’, ‘Stata’ and ‘SAS’ Files. R package version 2.3.1. https://CRAN.R-project.org/package=haven 11.Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01.