Potential influence of the non-vote voter in the 2019 Canada Election

Zhiang Chen

20/12/2020

Abstract

This study try to find the poential infulence of the one third Canadian that not vote in the 2019 canada election. If all Canadian who have the legal right to vote voted in the election, what will happened to the Liberal's Percentage of votes? In the study we use Multilevel logistic regression and post-stratification. And we find that the Liberal's votes rate slitly decrease if all Canadain voters vote.

Keywords

Election, Election Prediction, Voter Turnout, Election Prediction Model, Regression, Liberal Party

Introduction

In the 2019 Canadian election, the population of Canada is 35 million, but only 27 million of them registered as electors on the list, and only 18.35 million of them make their final vote on the election date. The voting rate is only 67.0%. The fact suggests that only 67% of Canadians, who have the legal right to vote, vote at the election day, and their decision finally comes to the new leadership of the country for the next four years. But what if all Canadian voters exercise their right to vote? The basic logic of democracy is every individual in the society chooses a public representative who has the most support of the people to make the public decision for the individual. The voting turnout is a representation of the degree of participation in this process, if the voter turnout rate is too low it is reasonable for the people to question if there exists a bias in this process.

In this study, we try to figure out the influence of the "silent 1/3" on the 2019 Canadian election result. We comparing the 2 prediction results from the model to come to the conclusion. In the first result, we only consider those who are likely to vote in the election to fit the logistic model and do the post-stratification to get the results. Then we get the hypothesis election result, we assume every Canadian who has the legal right to vote will vote, in other words giving the voter turnout of 100%. We considered all observations in our dataset and fit the logistic model and do the post-stratification, then we get the prediction result. By comparing the probability of the Liberal party's percentage of votes obtained in the election, we can conclude that the "silent 1/3" will slightly decrease the Liberal's percentage of votes obtained, considering the tight election results and the conservative party even have higher votes obtained rate, the "silent 1/3" might change the election result completely. In that case, we can say there may exist some bias in this real election results.

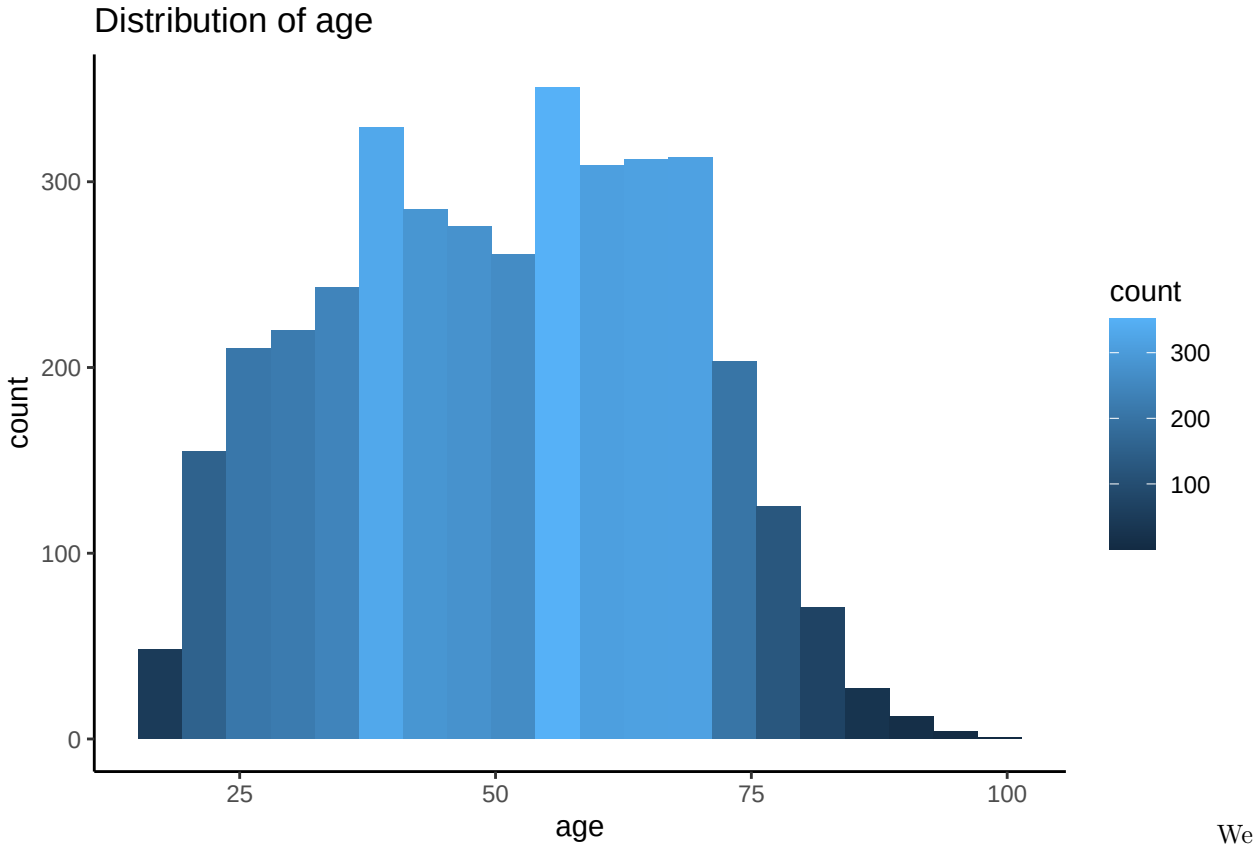
Data

In this paper we use 2 datasets, both of them are from the 2019 Canada election study. The researchers try to figure out the Canadian general public's opinion towards the election, the survey contains their answer to their question about their demographic characteristic, political opinion, and their attitudes towards a specific party. We will use this information to build a model to estimate the result under the condition that all Canadians will vote on the election date, and comparing the results to the estimated actual results.

We make an important assumption that in dataset "ces2019_phone", only people who are "certain to vote" and have a level of interest more than the average level of the interest of the federal election, or "already voted in the advanced poll" will vote certainly vote in the election. By doing this adjustment we can see the voting turnout is 65.95%, which is around the real turnout rate of 67%. We understand this assumption may have some bias, more detail will be discussed in the limitation part.

	c("66.88%", 65)
percent	66.88%
score	65

Another assumption we make is that only people who give a score of equal or more than 65 will finally vote for the Liberal since from the final result of the 2019 Canadian election, we know that 33.12% of the voters actually vote for the Liberal party, and the quantile is 65. We understand this assumption may cause some bias, more detail will be discussed in the limitation part.



We can see most of the people in our dataset is around 30-70, and there exist some people who are less than 25 and over 75 years old.

Var1	Freq
female	1629
male	2125
other	1

We can see there exists 1629 female, 2125 male, and 1 other gender in our dataset. The number of females is much less compared to the male.

Var1	Freq
fewer	857
more	1029
same	1869

We can see that most people are willing to accept the same number of immigrants which Canada is accepted right now, and the number who is willing to accept more immigrants is more than people who would like to accept fewer immigrants.

Var1	Freq
fewer	1031
more	992
same	1732

We can see that most people are willing to accept the same number of refugees which Canada is accepted right now, and the number who is willing to accept more refugees is less than people who would like to accept fewer refugees.

Var1	Freq
better	793
same	2137
worse	825

We can see most of the people have the same financial status under Justin Trudeau's administration, and the number who have worse financial status is higher than people who have better financial status.

Model

In this paper, we build a model to predict the Liberal party's voting rate base on people's age, gender, opinion on refugees and immigrants, and financial status. We use the multilevel regression and post-stratification to conclude. But we need to mention that the direct result from the model is not that important compared to the final results, the most important results are the difference between 2 estimate results, which can be used to show if every legal voter vote, what will the 2019 Canadian election results change.

We build the multilevel linear regression for this analysis. The dependent variable is "party" which represents whether each observation will vote for the Liberal party base on their attitude towards the Liberal party. Since we already know the election results, we know that the Liberal party wins 33.12% of the final election votes.

$$logit(\pi_i) = log(\frac{\pi_i}{1-\pi_i}) = \beta_0 + \beta_1 x_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_n x_{in}$$

In the model we will plug in our estimators as x_1 to x_n.

Observations	3755
Dependent variable	party
Type	Generalized linear model
Family	binomial
Link	logit

$\chi^2(9)$	585.91
Pseudo-R^2 (Cragg-Uhler)	0.20
Pseudo-R^2 (McFadden)	0.12
AIC	4338.03
BIC	4400.34

	Est.	S.E.	z val.	p
(Intercept)	-1.28	0.17	-7.47	0.00
age	0.00	0.00	1.25	0.21
sexmale	-0.42	0.07	-5.67	0.00
sexother	-12.45	196.97	-0.06	0.95
immigrantsmore	0.81	0.15	5.55	0.00
immigrantssame	0.61	0.13	4.64	0.00
refugeemore	1.27	0.14	9.27	0.00
refugeesame	0.95	0.12	7.83	0.00
financialsame	-0.58	0.09	-6.39	0.00
financialworse	-1.44	0.12	-11.61	0.00

Standard errors: MLE

We can see the p-value for age and sex which is other is not significant, for all the other variables is significant, thus we have strong evidence that the coefficient for all the other variables are not 0.

Post-Stratification

Table 6: Estimate results

	c("actual election result", 0.2829871)
full election result	actual election result
0.2721444	0.2829871

The real outcome of the election is 33.12%, which suggests that the Liberal party wins 33.12% of the vote by the Canadian voters. In our post-stratification, we get 28.2830% of votes which is similar to the election result. However, we can see if we consider every Canadian in this dataset, which implies that all Canadian who has legal right to vote will vote in the election, the Liberal's voting percentage will decrease by 1.08%, which suggests that if all Canadian vote, the Liberal may be won fewer votes. Considering that the conservative party already have a larger number of vote, this extra advantage may change the election results completely, thus we may argue that the election results are "biased".

Discussion

In this study we have many weaknesses and limitations that need to be discussed, the first one is the relatively small number of observations is the dataset that we use to build the model. In the 2019 Canada election study by phone data set, we only got around 4000 observations, which compared to the general population is relatively low, which may cause some bias.

Furthermore, we assume that only people who are satisfied "certain to vote" and have the level of interest of the election more than the average level will vote in the election, which may not true in the real world, since people may easily change their mind and make their vote. However we can see that there exist some people who declare that he or she will not vote, but they make their final vote in the election, on the other hands there also exists some people who declare them as they are definitely going to vote but not vote in the election data which may hedge the bias.

Moreover, we make another assumption that only people who give the score towards the Liberal party more than 65 will finally vote for the Liberal Party in the 2019 election. Even though in the "phone" dataset we can see this voting rate satisfied the actual voting rate, but in the "web" dataset might not be the case, thus this might cause some bias.

Also, in our model, we only considering people's age, gender, financial status, and political opinion towards refugees and immigrants, which may not be accurate, since we ignore a lot of information from the 2 questionnaire datasets. Some of them may be considered for a more complex logistic regression model.

For further study, we may focus on data, trying to use more convictive variables in the dataset to build a more complex dataset. For a more complex data set, the difference between the two fitted estimate results may be more convincing.

Reference

Voter Turnout at Federal Elections and Referendums. (n.d.). Retrieved from https://www.elections.ca/content.aspx?section=ele&dir=turn&document=index&lang=e#ftn1

R Core Team. 2020. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/. Stephenson, Allison Harell, Laura B., and Peter John Loewen. 2019. "Data:The 2019 Canadian Election Study – Phone Survey."

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." Journal of Open Source Software 4 (43): 1686.

Wickham, Hadley, Romain Fran?ois, Lionel Henry, and Kirill Müller. 2020. Dplyr: A Grammar of Data Manipulation. https://CRAN.R-project.org/package=dplyr.

Zhu, Hao. 2020. KableExtra: Construct Complex Table with 'Kable' and Pipe Syntax. https://CRAN. Rproject. org/package=kableExtra.