

2019 Canada Federal Election prediction when include all voters in

Pinlu Chen, 999940943

2020-12-09

Abstract

This paper uses the MRP method to predict the 2019 Canada Federal Election. The 2019 online survey data derivate the regression model to predict the resule, and then using census data to adjust the prediction error. The results shows 70% Canadian will vote Liberal party that 2019 Canada Federal Election, and the true result can reflect most of Canadians voting decision.

keywords: 2019 Canada Federal Election, MRP, prediction

Introduction

The US election has just passed, and news about cheating in the election is endless. Does the results of the election reflect the most wishes of all voters? It is a good question. Bailey pointed out that many people may give up election during the voting, and cause the result bias. In order to check it or make a prediciton for the future voting, statisticians widely use the MRP method to estimate the voting results.

In the 2019 federal election, Trudeau won 157 seats and was elected as prime minister, but he is still 13 seats short of forming the majority party. Minority governments must rely on other political parties' support to maintain power, which leads to less stability than that of majority governments. However, the voter turnout in this general election was only 67%, and nearly one-third of Canadians did not participate in the election. Therefore, it is meaningful to study the polls' results and help people choose The party representing the entire Canadians.

MRP method needs two databases, one reflecting people's willingness to vote and the other reflecting the entire Canadian electorate. This article obtains the first data based on the ces website online survey and then uses the gss data to reflect Canada's population distribution. In the following, this paper will introduce the data collection and variable selection. Then, based on the selected data set, MRP model analysis will be studied, and results and discussion will be in the last paragraph.

Data

Survey data

The dataset reflecting people's willingness to vote was collected from the Canadian election study website. Through internet research, 37,822 Canadian citizens and permanent residents, aged 18 or older. The survey database contains 620 variables. Only Part of the variables related to the election results were selected for analysis, such as age, gender, education level, work, income, family and other related factors. However, there are many missing data in the online survey; these data will lead to the deviation of the estimated results; the data needs to be cleaned up before regression analysis. The main tasks of cleaning include 1. Check the variable collection rate; if the collection rate is too low, discard the change; 2. Deal with abnormal points; 3. Code and classify the existing data. After data clean, age_group,sex,education,employment,bornin_canada,child,income,marital and Liberal were kepted in the dataset. The distributions of these variable are shown in Table 1.

Table 1: summary for survey data

Variabel	Overall = 18539
age_group	
less than 20	280
21 to 30	2124
31 to 40	3487
41 to 50	3193
51 to 60	3689
61 to 70	3865
above 70	1901
Sex	
Female	10154
Male	8385
education	
bachelor	6506
college	6206
Highschool	3311
Less than highschool	143
master or more	2373
employment	
full-time	8656
not full-time	9883
bornin_canada	
Yes	16146
No	2393
with child	
True	11818
False	6721
income	
less than \$25,000	2128
\$25,000 to \$49,999	3608
\$50,000 to \$74,999	3954
\$75,000 to \$99,999	3372
\$100,000 to \$ 124,999	1692
\$125,000 and more	3785
marital	
Married	8907
Living common-law	3032
Divorced	1449
Widowed	3193
Separated	724
Single, never married	3732
Liberal	
Yes	5508
No	13031

In Table 1, the age distribution of the people participating in the survey is relatively even. The participation rate of women is higher than that of men, and the participation rate of people with high education is higher than that of people with low education. Most people are born in Canada, and those with children are higher than those without children. , The income distribution is even, and the participation rate of people with good marriage status is high. The table data reflects the severe bias of the people participating in the survey, such as the low participation rate of people with low education and men's low participation rate. Only through

the survey results, we found that supporting the Liberal Party is less than 1/3. This is an extraordinary result. After all, the actual election result is only 13 seats short of the Liberal Party.

Gss data

The government initiated Canada's General Social Survey (GSS) to collect data on relevant social trends to detect Canadians' current social habits and guide policy formulation. The samplers targeted by GSS are Canadians over 15 years old. Generally, online questionnaires and telephone collection are used to randomly sample samples to collect data. Therefore, this method is more representative and can reflect the distribution of Canadian population. Therefore, it can be used as post-stratification data to adjust the deviation of the election rate estimated by the Canadian election study website survey data. Figure 1 shows the distribution of Canadians' age and education level. The pattern indicates that the low-educated population is close to 10%. However, in the previous survey, the proportion is less than 1% in the survey data.

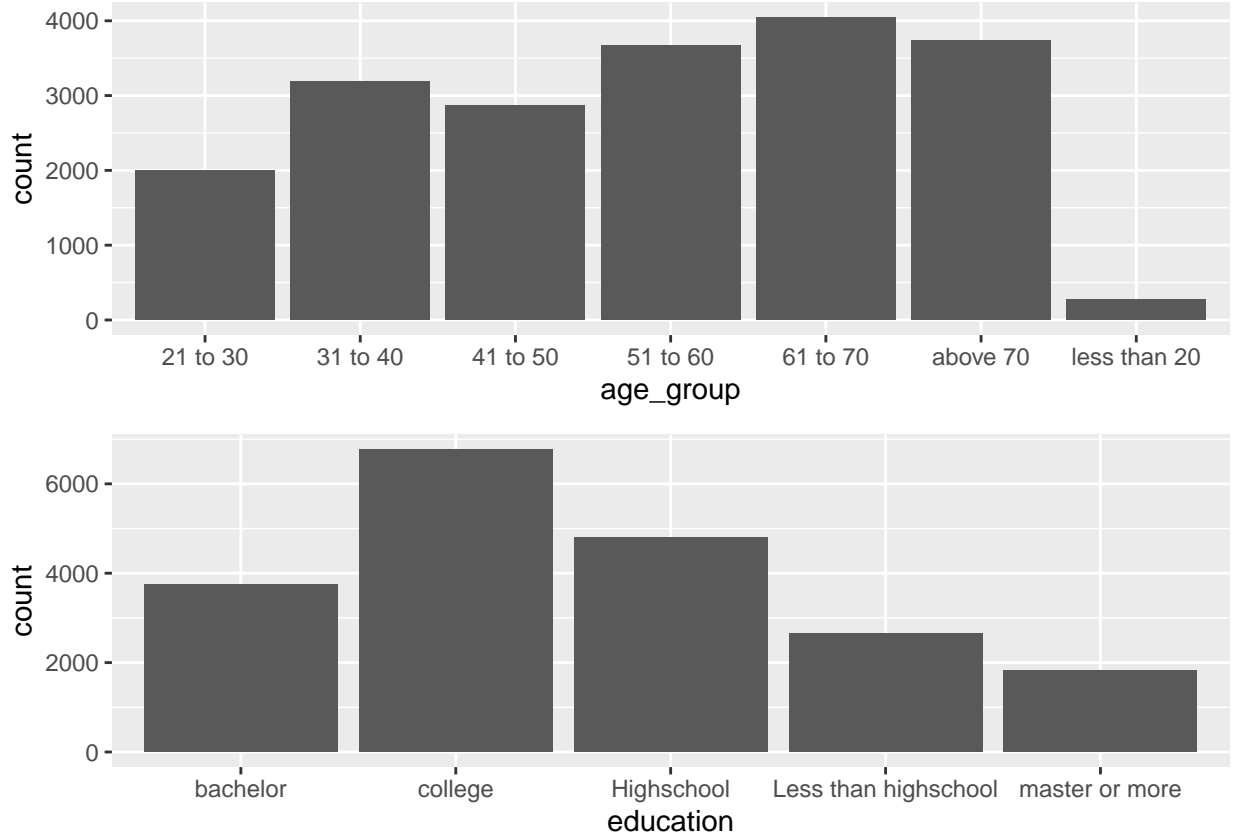


Figure 1: Distribution of Age and education in GSS

Model

I will be using a logit regression model to model the proportion of voters who will vote for Liberal party. The formula for the candidate model is:

$$\text{logit}(p) = \beta_0 + \beta_1 x_{age} + \beta_2 x_{sex} + \beta_3 x_{edu} + \beta_4 x_{emp} + \beta_5 x_{bornin} + \beta_6 x_{child} + \beta_7 x_{income} + \beta_8 x_{marital} + \epsilon$$

Where p is the probability voting liberal party. The logit function is in the form:

$$\eta = \text{logit}(p) = \log\left(\frac{p}{1-p}\right)$$

The glm regression will estimate the value of logit function, and then we can use the $\hat{p} = \frac{exp^{\eta}}{1+xp^{\eta}}$ to get the probability. Under the backward model selection by using AIC, only age_group, sex, education, bornin_canada and child are remained in the model.

After obtaining the estimated model, we will clean the related variables in gss dataset and then divide them into cells according to the selected variables. Finally, we obtain the census data use it to estimate the probability of the Liberal Party winning when the entire Canadian population participates in the election based on the logit model.

Results

The result of the estimated model is shown in Table 2. Through the analysis of the model results, with the increase of age, the support rate of the Liberal Party has a downward trend; Canadians with high education level have a lower support rate of the Liberal Party. For example, compared with the bachelor, highschool and college, the estimated parameters are all positive, indicating that highschool And the college crowd are more willing to choose the Liberal Party. In addition, people born in Canada and people with children also have a higher support rate for the Liberal Party. Under the MRP process, we get the voting rate is 70%, much higher than the one shown in the Canadian election study website.

Table 2 Final logit model of voting liberal

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.53	0.07	7.87	0.00
age_group31 to 40	-0.17	0.06	-2.78	0.01
age_group41 to 50	-0.13	0.06	-1.99	0.05
age_group51 to 60	-0.21	0.06	-3.33	0.00
age_group61 to 70	-0.29	0.06	-4.64	0.00
age_groupabove 70	-0.41	0.07	-5.66	0.00
age_groupless than 20	-0.18	0.14	-1.26	0.21
sexMale	0.05	0.03	1.44	0.15
educationcollege	0.34	0.04	8.69	0.00
educationHighschool	0.53	0.05	10.63	0.00
educationLess than highschool	-0.03	0.18	-0.18	0.86
educationmaster or more	-0.15	0.05	-2.88	0.00
bornin_canadaYes	0.25	0.05	5.36	0.00
childTRUE	0.19	0.04	5.20	0.00

Discussion

Through MRP analysis, we got shocking results, 70% of Canadians will choose the Liberal Party. This result is reasonable because the participation rate in this general election is only 67%, and nearly half of those who participated in the election chose the Liberal Party. In the above logit regression analysis, we found that low-educated, young people born locally and with children are more willing to choose the Liberal Party. For example, when keeping other variables constant, the log odds of Canadians who were born only locally who chose the Liberal Party were 0.25 higher than those who were born in Canada. Although Canada is an immigrant country, the local-born population accounts for a large proportion and is still the mainstream population. Besides, families with children are more willing to choose the Liberal Party and favored by people with low education. The Liberal Party will therefore receive a high support rate. Although the 2019 federal election will establish a government in the form of a minority party, it still reflects what Canadians think. In future elections, the government should encourage voters to vote.

Weaknesses

This article does not use census data as the post-stratification dataset but uses GSS data, which may introduce estimation errors. In the GSS data, there are some variables with a very low response rate. Although this variable may be related to the election results, the low participation rate is not conducive to the estimation's accuracy. In the subsequent research, we can use the census database for analysis to obtain more accurate results.

References

Stephenson, Laura B., Allison Harell, Daniel Rubenson and Peter John Loewen. The 2019 Canadian Election Study – Online Collection. [dataset]

Election Canada (n.d.). Retrieved December 20, 2020, from <https://www.elections.ca/content.aspx?section=ele>