

# Prediction on the 2019 Canada Federal Election based on MRP

Liu, Zhitong

2020-12-18

## Abstract

This study predicts the result of 2019 Canada Federal Election based on MRP method. Using the survey data to estimate the regression model, and then according to the census data to adjust the prediction. The results shows that both the Liberal Party and the Conservative Party are not enough to get the majority of votes.

**keywords:** Voting prediction, MRP, Canada Federal Election

## Introduction

Predicting election results is one of the hottest applications in statistics. The anticipated results may influence future elections. However, due to sampling error, the survey data collected from some particular website cannot represent the whole population. The result has some bias. How to reduce estimated bias is an essential research issue in statistics. MRP model is one of the methods used to solve this problem.

According to Jack Bailey's idea, some voters will abandon the election during the election process, and this would cause the final election results may not reflect the population. The participation rate in the 2019 Canadian Federal Election is only 67%. The election's outcome was that the Liberal Party won the general election, but there were no more than half of the seats, and it could only be a minority party. The low Voter turnout may be the big concern in the voting.

The CES website did an online survey to document Canadians' attitudes during and after the 2019 election. However, the surveyors on this website may not represent the whole population. Thus, we need another dataset to adjust the estimation. In the following, the multilevel regression and poststratification (MRP) method is used to estimate the voting outcome, including all the voters. we can use the calculated result to check the 2019 Canada Federal Election result. In the first section, survey data and gss data will be introduced , then how the MRP model will be built. The results of the MRP are listed in section 3. In the end, a discussion on the results and potential weaknesses will be given.

## Data

The CES online survey's original data contains 37822 observations and 620 variables, but there are too many NAs in these data. Before the follow-up analysis, we must clean these data first, according to the data response rate and importance. After cleaning, we finally select 31115 observations and nine variables. These data are summarized in Table 1. Because the data needs to be divided into cells in the MRP method, the digital variable age is divided into ten years old and divided into multiple groups in the data cleaning process, and individuals younger than 18 years old are filtered out. We believe that age, gender, education level, work status, and family situation will all affect the election.

Through the summary table, we found that there is a serious imbalance in the data. For example, the number of women is 7000 more than the number of men, and only about 1% of the population with low education. Therefore, it is unreasonable to judge the general election situation only through the survey results, and the model needs to be revised with census data to obtain more accurate results. However, census data is difficult to obtain. This article uses GSS data instead of census data to describe the distribution of the Canadian population. The GSS data is officially collected and published and has a high reference value. In

selecting survey variables, we must also consider the corresponding variables in the gss data. Figure 1 shows the distribution of age and education in the GSS data. The figure shows that the low-educated population accounts for nearly 10%. However, only 1% of the CES online survey data is a serious shortage.

**Table 1: Survey data**

Variabes	Total = 31115
<b>age</b>	
18 to 20	1221
21 to 30	5247
31 to 40	7112
41 to 50	6221
51 to 60	7121
61 to 70	7347
above 70	3553
<b>Male</b>	
FALSE	22271
TRUE	15551
<b>education</b>	
bachelor	12908
college	12096
Highschool	7514
Less than highschool	423
master	4782
<b>full-time</b>	
FALSE	21698
TRUE	16124
<b>born_in_canada</b>	
FALSE	6266
TRUE	31556
<b>has_child</b>	
True	22530
False	14904
<b>marital</b>	
Married	17230
Living common-law	6070
Divorced	2722
Widowed	1395
Separated	1374
Single, never married	8458
<b>Liberal_party</b>	
TRUE	8949
FALSE	22615
<b>Conservative_part</b>	
TRUE	8713
FALSE	22851

## Model

According to the dummy type of output variable, we use logistic regression model to analyze it. The regression model formula is as follows:

$$\text{logit}(p) = X\beta + \epsilon$$

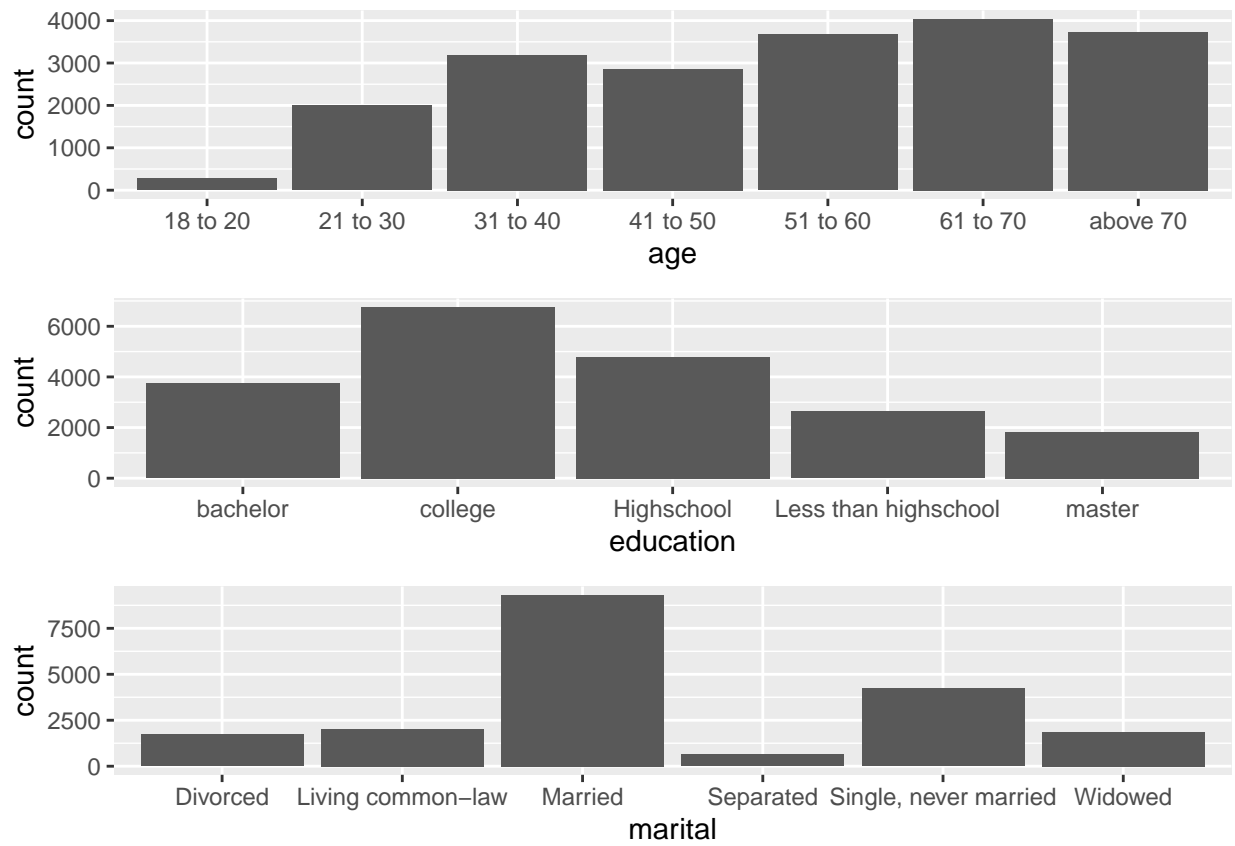


Figure 1: Distribution of age, education and marital in Canada

where X is the predictor variables, p is the vote rate. However, the estimated result is just log odd, we need to do further transformation to obtain the success probability. The transformation formula is  $\hat{p} = \frac{e^x \beta}{1 + e^x \beta}$ . After get the optimal model by AIC, the estimated probability will be adjusted according to the GSS data using Post stratification method.

## Results

We first estimate the election rate of the Liberal Party. Through the reverse selection method, we get the best estimation model. The results are shown in Table 2. The results show that age, education level, place of birth, and children are significant. Among them, families born in Canada and with children have a lower election rate for the Liberal Party. At the same time, a bachelor or more with higher education is more willing to support the Liberal Party. The supproting rate of Corresponding results of the Conservative Party in Table 3 show that age, gender, education, place of birth, children, and marital status are all significant. Families with children are more willing to support the Conservative Party. Low-educated and male support for the Conservative Party is higher, while adults also support the Conservative Party. The Liberal Party's support rate is 27.8%, and the Conservative Party's support rate is 29.5%, which is higher than the Liberal Party.

**Table 2 results of Liberal\_party**

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-0.36	0.08	-4.31	0.00
age21 to 30	-0.26	0.08	-3.15	0.00
age31 to 40	-0.09	0.08	-1.11	0.27
age41 to 50	-0.15	0.08	-1.79	0.07
age51 to 60	-0.09	0.08	-1.13	0.26
age61 to 70	0.03	0.08	0.37	0.71
ageabove 70	0.12	0.09	1.41	0.16
educationcollege	-0.36	0.03	-11.69	0.00
educationHighschool	-0.52	0.04	-13.72	0.00
educationLess than highschool	-0.13	0.13	-1.01	0.31
educationmaster	0.13	0.04	3.31	0.00
born_in_canadaTRUE	-0.23	0.04	-6.56	0.00
has_childTRUE	-0.17	0.03	-5.93	0.00

**Table 3 Results of Conservative\_part**

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-1.73	0.12	-14.87	0.00
age21 to 30	0.33	0.10	3.27	0.00
age31 to 40	0.28	0.10	2.76	0.01
age41 to 50	0.37	0.10	3.66	0.00
age51 to 60	0.37	0.10	3.67	0.00
age61 to 70	0.31	0.10	3.00	0.00
ageabove 70	0.39	0.11	3.61	0.00
MaleTRUE	0.42	0.03	15.78	0.00
educationcollege	0.27	0.03	8.64	0.00
educationHighschool	0.37	0.04	9.90	0.00
educationLess than highschool	-0.02	0.14	-0.14	0.89
educationmaster	-0.24	0.05	-5.27	0.00
born_in_canadaTRUE	-0.15	0.04	-4.17	0.00
has_childTRUE	0.15	0.03	4.53	0.00
maritalLiving common-law	-0.10	0.06	-1.71	0.09
maritalMarried	0.42	0.05	8.03	0.00

	Estimate	Std. Error	z value	Pr(> z )
maritalSeparated	-0.24	0.09	-2.75	0.01
maritalSingle, never married	-0.22	0.06	-3.48	0.00
maritalWidowed	0.17	0.08	2.06	0.04

## Discussion

The MRP method is used to estimate the support rate of the Liberal Party and the Conservative Party in 2019 Federal elections. The results show that the support rate of the two parties is not enough to exceed 50%. Therefore, in the 2019 federal election, the Liberal Party won the election with a slight advantage, and the formation of a minority government reflects certain Canadian current conditions. Comparing the regression results, we find that the Liberal Party and the Conservative Party are quite different in some individuals. For example, the Conservative Party favours middle-aged people and people with children, while young people and highly educated people favour the Liberal Party. High approval rate. However, in some respects, the policies of these two parties are not pleased. For example, born in Canada has negative coefficients.

## Weaknesses

Although here uses GSS data to correct the deviations in the results of survey estimation, due to the large amount of missing data in GSS data, GSS can only reflect Canadians' tendencies to a certain extent. However, we must realize that it is still consistent The census data is biased. If the census data crowded MRP estimates are available, the results will be more credible.

## References

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