

Prediction of the Overall Popular Vote of the Next Canadian Federal Election

STA304 - Assignment 2

Mengyang Liu, Pengying Li, Bowei Cheng, Shan Huang

November 24, 2022

Introduction

It has been a year since the last federal election in 2021, in the 2021 federal election Justin Trudeau and his liberal party won his third term as the prime minister. The importance of elections is self-evident. In a democracy, people have the right to vote for the leader who is best for them, and the result of the election will decide the direction of the country's development. Sometimes the result of the federal election will have a significant influence on many parts of society such as the financial market. Thus, it is important to predict the result of the federal election.

However, the result of the election is hard to predict. Because different people have different preferences for voting. But there are some research shows that there are some factors that may affect people's voting choices such as age, family income, living province, and so on (Stephenson et al., 2021). While different parties have different opinions dealing with social issues. For household income, the different parties have different opinions on income inequity. For example, the green party intends to introduce a wealth tax in Ontario because they believe it will reduce income inequality. However, it may hurt the benefit of those high-income families. Thus, those high-income families may be less likely to vote green party (Welch & Hibbing, 1992). On the other hand, the green party also intends to increase the rate of Ontario Works (OW) and Ontario Disability Support Program (ODSP) which may help those low-income families and lead these low-income families to vote for them (Armstrong et al., 2022). Here we can see that these factors may be helpful for predicting the result of the federal election.

In this assignment, our research question is which party most likely won the 2025 election. The reason why we choose this question is that answering this question, it can provide society with an expectation of how their lives will be. Because different parties will have different policies and actions in facing different events.

We will be using two data sets, one is survey data called Canadian Election Survey (CES) and the other one is census data called General Social Survey (GSS). The CES was conducted in 2019 and it was used to get Canadians' attitudes and opinions for the 2019 federal election. There are 20602 observations and 61 variables. The General Social Survey (GSS), it is established in 1985. It is used to get data about the social trend including people's living conditions and how the information policy influences society. More details about these two data sets such as important variables will be provided in the data section.

The logistics regression model is used to predict the winning probability of each party. With this model, we will use gender, age, family income, current living province, education level, home language, and life feeling these 7 variables as predictors to predict the winning probability of the popular parties in Canada and use the model to perform post-stratification and identify which party is most likely to win the 2025 federal election.

Data Cleaning

For survey data (CES2019 phone data), we just did three kinds of manipulations, which are creating new variables, selecting those new variables that we will use, and removing any missing values. First, we selected

8 variables which are “q2, q3, q70, q4, q61, q67, q6, q11”. Then we created 13 new variables. We used q2 to generate “age”, used q3 to create “gender” by looking at the codebook to see what is the meaning of each code value in that variable. For example, the code value “1” is male, and “2” is female. We continued to use that method to create a “current living province” by using q4. For creating other variables, we not only needed to just look at the codebook, but we also needed to classify the code values into different groups. For example, for “family income”, we used q70, and when q70 is “1” or “2”, we labeled it as “low income”, when q70 is “3”, “4”, or “5”, we labeled it as “medium income” and when q70 is “6”, “7”, “8”, we labeled it as “high income”. By repeating the same process, we created “education level” by using q61, and if q61 is “1” “2” “3” “4” or “5”, we classified it as “less than high school and high school” and if q61 is “6” “7” “8” or “9”, we classified it as “college and university”, otherwise, we labeled it as “master and Ph.D. degree”. Repeating the process, we created a “home language” by using q67 and we labeled it as “English only(q67 is 1)”, “French only(4)”, and “Don’t have the official language of Canada(otherwise)”. And then we used q6 to create “life feelings”, and classified them as “very satisfied(q6 is 1)”, “somehow satisfied(2, 3)”, and “very dissatisfied(4)”. And finally, used q11 to generate “vote for Liberal”, “vote for Conservative”, “vote for NDP”, “vote for QB”, “vote for Green party”, and “vote for People’s party”. The last six variables are binary variables. For example, “vote for Liberal” is 1 if people voted for Liberal, otherwise, it is 0. Lastly, we selected those new variables and removed any missing values.

For cleaning the census data (GSS data), we also did the same three steps before, creating new variables, selecting those new variables that we will use, and finally removing any missing values. There are 7 new variables in total and all those variables’ names are the same as the first 7 new variables in the survey data. For “family income”, “education level”, “home language”, and “life feelings”, we also classified each variable into several groups. For “family income”, we used the variable “income family” to generate. And if the income is “Less than \$25,000”, it is “low income”, if it is “\$25,000 to \$49,999” or “\$50,000 to \$74,999” or “\$75,000 to \$99,999”, we labeled “medium income” and if it is “\$100,000 to \$ 124,999” or “\$125,000 and more”, we labeled “high income”. For “education level”, we used “education” to generate, and if education is “Less than high school diploma or its equivalent” or “High school diploma or a high school equivalency certificate”, we labeled “Less than high school and high school”, if the education is “Trade certificate or diploma”, “College, CEGEP or other non-university certificate or di. . .”, “University certificate or diploma below the bachelor’s level”, Bachelor’s degree (e.g. B.A., B.Sc., LL.B.), we labeled “college and university”, otherwise, we labeled “master and Ph.D. degree”. For “home language”, we used “language home” in data to generate. If “language home” is “English” or “English and non-official language”, we labeled “English only”, if it is “French” and “French and non-official language”, we labeled “French only”, if it is “English and French”, we labeled “Both official languages” and if it is “Multiple non-official languages” or “Non-official languages”, we labeled “Don’t have the official language of Canada”. Last one is about “life feelings”, we used “feelings life” to generate. And we label it is “very satisfied(7, 8, 9, 10)”, “somehow satisfied(4, 5, 6)” and “very dissatisfied(1, 2, 3)”. After we created those new variables, we selected them and removed any missing values in this data set.

Introducing Important Variables

The important variables are age, education level, home language, and family income. Based on our research, those factors play more important roles in the voting opinion of the political party(Uppal & LaRochelle-Côté, 2015). Age is just the age of people when they voted in 2019, education level just describes until 2019, what is the highest education level they achieved. Home language could reflect the culture to some degree, and it just illustrates the first language people learned the last one is family income, which just describes the income category that each individual’s family gained a year.

We will use the cleaned survey data to build a logistic regression model for each party. In those models, the response variables are the votes for each party, such as the “vote for Liberal” or “vote for Conservative”. The remaining 7 variables are predictors, which are age, gender, family income, current living province, home language, life feelings, and education level. We selected those variables since all of them could affect people’s decisions of voting(Uppal & LaRochelle-Côté, 2015). For each party, we want to use those variables to predict the proportion of Canadians who vote for that party to see which party is the most popular in Canada.

Graphs and Tables

We did some research and we found that the Liberal and Conservative parties always gain a high number of votes among Canadians(Canadian election results by party 1867 to 2021), so we drew the graphs to see how the predictors will affect people's decisions about voting for the Liberal or Conservative parties.

Table 1: Statistaics of the age of all voters in the survey data

Min	Q1	Median	Q3	Max	IQR	Mean	SD
18	38	55	68	100	30	52.4281	19.03213

In our survey data, the mean age is 52.4, the minimum age is 18 and the maximum age is 100. About 75% of people who completed the survey were aged below 68.

Fig.1 Barplot of Voting the Liberal Party in Different Provinces

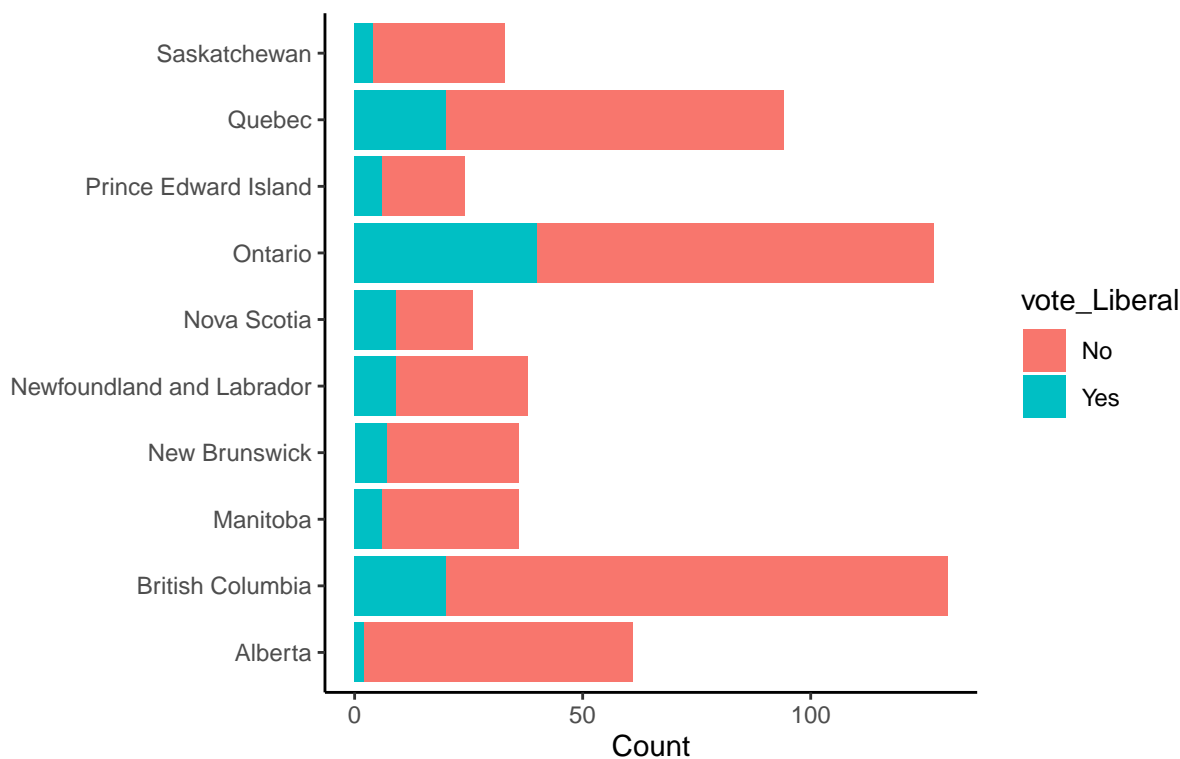


Table 2: Statistaics of voting for Liberal Party in Different Provinces

Province	Number of people vote for Liberal Party
Alberta	2
British Columbia	20
Manitoba	6
New Brunswick	7
Newfoundland and Labrador	9
Nova Scotia	9
Ontario	40
Prince Edward Island	6

Province	Number of people vote for Liberal Party
Quebec	20
Saskatchewan	4

Figure.1 shows the number of people who voted Liberal in each province in Canada. We could see that among the people who voted Liberal, people in Ontario weigh the largest proportion, and people in Alberta weigh the least proportion. There are the about same amount of people in Prince Edward Island, New Brunswick, and Manitoba who voted for the Liberal party. Table 2 provides a clear summary for the number of the Liberal Party supporter in different provinces.

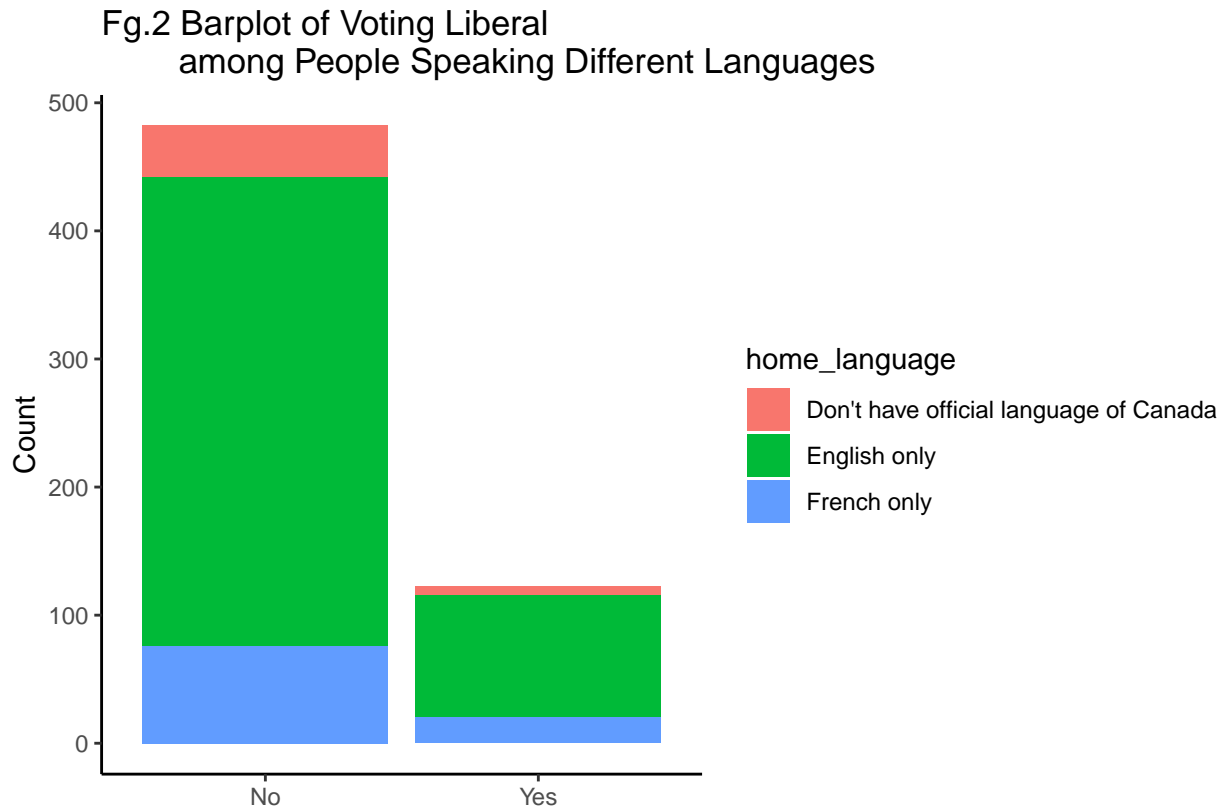


Table 3: Statistaics of voting for Liberal Party with different language

Language	Number of people vote for Liberal Party
Don't have official language of Canada	7
English only	96
French only	20

Figure.2 wants to illustrate how culture will affect people's decisions about voting Liberal. Culture influences a lot about voting decisions(Voting behavior in Canada). We can see that people speaking English would have a high probability to vote Liberal. But we also can see those people also have a high probability of not voting Liberal. This may be due to the problem of classifying the language spoken when we cleaned the data. If people can speak English, we just put them into the "English" category. So, there is a large proportion of

English-speaking people in the data. Table 3 provides a clear summary for the number of the Liberal Party supporter with different home languages.

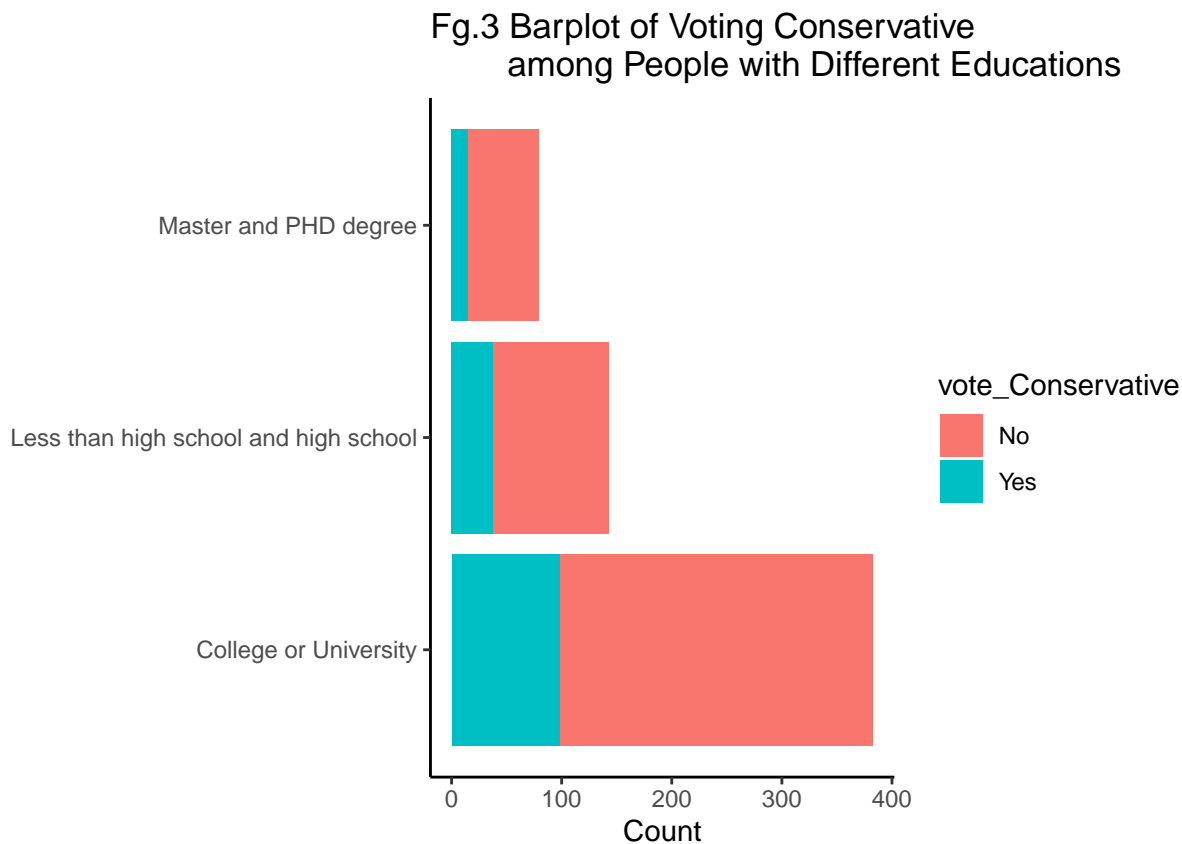


Table 4: Statistaics of voting for Conservative Party with different education level

Education level	Number of people vote for conservative Party
College or University	98
Less than high school and high school	38
Master and PHD degree	15

Figure.3 shows how the education level of Canadians affects the vote opinion about the Conservative party. We can see that the probability of voting for the Conservative party among college and university people is higher than the other education level. Table 4 provides a clear summary for the number of the Conservative Party supporter in different education levels.

Fig.4 Barplot of Home Languages

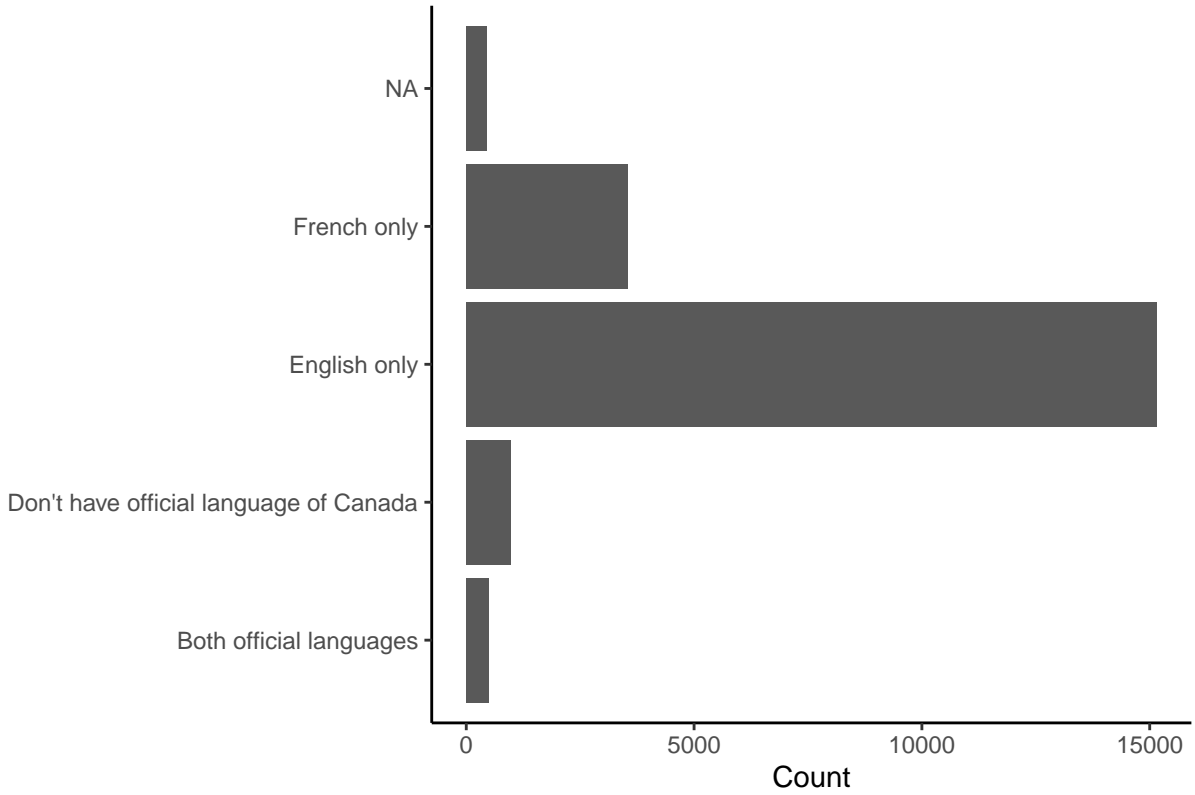


Figure.4 is about census data. This is the only variable that we have some problems with matching to survey data. To match this variable to survey data, we delete the category “Both official languages” and any missing values. We think it would be fine to delete them since there is less data related to these two categories by looking at the bar graph.

Methods

Logistic regression model

A logistic regression model will be used to predict the support rate of popular parties in Canada based on the dataset collected from the 2019 Canadian election study survey. As the outcome of our prediction should be either people will vote for one specific party or people will not do so, it is considered as a binary response. Therefore, a logistic regression model is suited to do such estimation.

Seven predictors will be used in our logistic regression model which include age, gender, current living province, home language, education level, family income and feeling of life. Age is the only numerical variable, and the other ones are all categorical predictors. The response variable is whether people will vote for a party or not, and there are six parties that we consider as popular in Canada. Therefore, we build six logistic models for each of them with the same predictors.

The general logistic regression model that we use is:

$$\begin{aligned} \log\left(\frac{p}{1-p}\right) = & \beta_0 + \beta_1 x_{age} + \beta_2 x_{Male} + \beta_3 x_{LiveBC} + \beta_4 x_{LiveMT} + \beta_5 x_{LiveNB} + \beta_6 x_{LiveNL} + \beta_7 x_{LiveNS} \\ & + \beta_8 x_{LiveON} + \beta_9 x_{LivePE} + \beta_{10} x_{LiveQB} + \beta_{11} x_{LiveSK} + \beta_{12} x_{Lowincome} + \beta_{13} x_{medincome} \\ & + \beta_{14} x_{Highschool} + \beta_{15} x_{MasterPHD} + \beta_{16} x_{English} + \beta_{17} x_{French} + \beta_{18} x_{dissatisfied} + \beta_{19} x_{satisfied} \end{aligned}$$

Where:

- p represents the probability of people will vote for our interested party.
- β_0 is the intercept of the model, and it is the log of odds of voting for our interested party when it is female from Alberta with high income, graduate from college or university, speaking non-official language of Canada and feel somehow satisfied to their life.
- β_1 is the slope of age, and for every one-unit increase in age there will be a β_1 increase log odds of voting for our interested party.
- β_2 is the coefficient of variable with gender of male.
- β_3 is the coefficient of variable with living province of British Columbia.
- β_4 is the coefficient of variable with living province of Manitoba.
- β_5 is the coefficient of variable with living province of New Brunswick.
- β_6 is the coefficient of variable with living province of Newfoundland and Labrador.
- β_7 is the coefficient of variable with living province of Nova Scotia.
- β_8 is the coefficient of variable with living province of Ontario.
- β_9 is the coefficient of variable with living province of Prince Edward Island.
- β_{10} is the coefficient of variable with living province of Quebec.
- β_{11} is the coefficient of variable with living province of Saskatchewan.
- β_{12} is the coefficient of variable with low family income.
- β_{13} is the coefficient of variable with medium family income.
- β_{14} is the coefficient of variable with Less than high school and high school education level.
- β_{15} is the coefficient of variable with Master and PHD degree education level.
- β_{16} is the coefficient of variable with home language only English.
- β_{17} is the coefficient of variable with home language only French.
- β_{18} is the coefficient of variable with very dissatisfied life feeling.
- β_{19} is the coefficient of variable with very satisfied life feeling.

For logistic regression we need to check these 4 assumptions:

1. Whether outcome is binary: To check assumption 1, we construct scatterplots of our response variable for the six popular parties in Canada that we are interested in. By observation from figure.5 in the appendix part, they are all binary.
2. Linearity in the logit for continuous variables We simply check this assumption using the Box-Tidwell test (STA304 Week 6 code session) By constructing a hypothesis test having H_0 : linearity between continuous variable and log-odds vs. H_A : non-linearity between continuous variable and log-odds. As there is only age as a numerical variable, we simply have a quick check on it. It seems all p-values are not significant by age, so we fail to reject H_0 : and say the assumption is not violated.
3. It should be lack of strongly influential outliers By checking the cook's distance from figure.6 in the appendix which is about our six interest variables, it shows that there is no strong influential outliers
4. There is no multicollinearity We use the variance inflation factors to check if this assumption is violated. There is no VIF value that exceeds 5 which indicates there is no problematic amount of collinearity

Post-Stratification

After constructing the logistic regression models for each party, we also have to perform a post-stratification analysis. It is done by making different cells with different ages, gender, living province, home language, education level, life feelings and family incomes. Then, we match each cell to the census data to reweight the proportion based on the population. Since my models are using the same predictors, I only need to match the same thing but with different models. The model will predict the probability of each cell. A general idea is to change each cell with their respective population weight, then we sum them all and divide the total population size. The post-stratification idea could be performed as this mathematical form:

$$\hat{y}^{PS} = \frac{\sum N_j \hat{y}_j}{N}$$

Where:

- \hat{y}^{PS} represents the estimation after using the post-stratification method.
- N_j is the population size of each stratum j.
- \hat{y}_j represents the sample mean for each stratum j.
- N is the total population size.

It is important to do so because our estimation is based on the survey data which may not performed well when the census dataset has different distribution on our predictors. Therefore, we should use the known population information to make our sample-based model prediction closer to the target population.

All analysis for this report was programmed using R version 4.0.2.

Results

Logistic Regression Model for the Liberal Party

$$\begin{aligned} \log\left(\frac{p}{1-p}\right) = & -4.240716 + 0.014001x_{age} - 0.059066x_{Male} + 1.536470x_{LiveBC} + 1.724456x_{LiveMT} \\ & + 2.033971x_{LiveNB} + 2.030563x_{LiveNL} + 2.734333x_{LiveNS} + 2.588633x_{LiveON} + 2.248234x_{LivePE} + 2.259589x_{LiveQB} \\ & + 1.584160x_{LiveSK} - 0.249441x_{Lowincome} - 0.087924x_{medincome} - 0.309670x_{Highschool} - 0.311215x_{MasterPHD} \\ & + 0.347780x_{English} + 0.202580x_{French} - 1.787547x_{dissatisfied} + 0.705319x_{satisfied} \end{aligned}$$

Logistic Regression Model for the Conservative Party

$$\begin{aligned} \log\left(\frac{p}{1-p}\right) = & -0.034288 + 0.015954x_{age} + 0.369978x_{Male} - 1.117324x_{LiveBC} - 0.926920x_{LiveMT} \\ & - 1.612914x_{LiveNB} - 2.747911x_{LiveNL} - 1.590368x_{LiveNS} - 1.786551x_{LiveON} - 1.452267x_{LivePE} - 2.360317x_{LiveQB} \\ & - 1.189701x_{LiveSK} - 0.190448x_{Lowincome} - 0.070502x_{medincome} - 0.028695x_{Highschool} - 0.553081x_{MasterPHD} \\ & - 0.706751x_{English} - 0.907791x_{French} + 0.502419x_{dissatisfied} + 0.140448x_{satisfied} \end{aligned}$$

Logistic Regression Model for the NDP Party

$$\begin{aligned} \log\left(\frac{p}{1-p}\right) = & -1.376057 - 0.048416x_{age} + 0.167639x_{Male} + 0.587200x_{LiveBC} + 1.153492x_{LiveMT} \\ & - 0.249050x_{LiveNB} + 1.002244x_{LiveNL} + 0.414599x_{LiveNS} + 0.436717x_{LiveON} - 14.802874x_{LivePE} - 0.413068x_{LiveQB} \\ & + 1.040289x_{LiveSK} - 0.404544x_{Lowincome} - 0.104911x_{medincome} + 0.015721x_{Highschool} + 0.834642x_{MasterPHD} \\ & + 0.896352x_{English} - 0.339360x_{French} + 0.815630x_{dissatisfied} - 0.603296x_{satisfied} \end{aligned}$$

Logistic Regression Model for the Quebec Party

$$\begin{aligned} \log\left(\frac{p}{1-p}\right) = & (-4.329e+01) + (1.433e-02)x_{age} + (1.033e+00)x_{Male} + (-1.328e+01)x_{LiveBC} + (6.558e-02)x_{LiveMT} \\ & + (-1.664e+01)x_{LiveNB} + (-2.008e-01)x_{LiveNL} + (2.056e-01)x_{LiveNS} + (-1.283e+01)x_{LiveON} + (-5.175e-02)x_{LivePE} \\ & + (3.137e+00)x_{LiveQB} + (-5.884e-02)x_{LiveSK} + (1.775e+01)x_{Lowincome} + (1.753e+01)x_{medincome} \\ & + (8.177e-01)x_{English} + (1.937e+01)x_{French} + (-1.993e+01)x_{dissatisfied} + (-1.849e+01)x_{satisfied} \\ & + (-6.303e-01)x_{Highschool} + (-1.953e+01)x_{MasterPHD} \end{aligned}$$

Logistic Regression Model for the Green Party

$$\begin{aligned} \log\left(\frac{p}{1-p}\right) = & (-3.315e+00) + (-2.621e-02)x_{age} + (2.448e-01)x_{Male} + (1.353e+00)x_{LiveBC} + (1.368e+00)x_{LiveMT} \\ & + (2.964e+00)x_{LiveNB} + (1.339e+00)x_{LiveNL} + (2.127e+00)x_{LiveNS} + (1.531e+00)x_{LiveON} + (1.729e+00)x_{LivePE} \\ & + (6.667e-01)x_{LiveQB} + (-1.449e+01)x_{LiveSK} + (5.922e-01)x_{Lowincome} + (4.910e-01)x_{medincome} \\ & + (1.653e-01)x_{Highschool} + (4.931e-01)x_{MasterPHD} + (-1.233e-01)x_{English} + (-2.375e-01)x_{French} \\ & + (1.921e-02)x_{dissatisfied} + (-1.179e+00)x_{satisfied} \end{aligned}$$

Logistic Regression Model for the People Party

$$\begin{aligned} \log\left(\frac{p}{1-p}\right) = & -4.07393 - 0.01040x_{age} + 0.49283x_{Male} - 0.75550x_{LiveBC} - 17.73776x_{LiveMT} \\ & - 18.03481x_{LiveNB} - 17.97873x_{LiveNL} - 17.92462x_{LiveNS} + 0.06287x_{LiveON} + 0.27404x_{LivePE} - 0.98643x_{LiveQB} \\ & + 0.20214x_{LiveSK} + 1.85749x_{Lowincome} + 1.39072x_{medincome} - 0.93969x_{Highschool} + 0.89524x_{MasterPHD} \\ & - 0.13480x_{English} - 0.44095x_{French} - 17.76578x_{dissatisfied} - 17.49295x_{satisfied} \end{aligned}$$

Based on the previous research, we got six logistics regression models that are corresponding to each party in the federal election. Then, we shall perform the post-stratification based on these logistics regression models.

Table 5: Estimate Result

Liberal	Conservative	NDP	QB	Green	People
0.3313156	0.242909	0.0514529	0.0012912	0.0260282	0.002226

By using the post-stratification, we got the estimate of the winning probability for each party in the 2025 federal election. As shown in Table 5, we can see the voting probability for each party. the voting probability of the Liberal party is 33.13156%, for the Conservative party is 24.290%, for New Democratic Party is 5.14529%, for Bloc Québécois is 0.12912%, for the Green party is 2.60282%, and for people party of Canada only stands for 0.2226%. By analyzing these values, we can see that the Liberal party and Conservative will stand for the most voting rate in the 2025 federal election. The Liberal party is expected to have the highest voting rate which means that the liberal party will be most likely to win the 2025 federal election.

Assessing the validity of these, we can compare the result of our research with the previous federal election result and check the p-value of each model. In our model, we saw that there are only two models having most of the variables with p-values smaller than 0.05, which are the Conservation party and the Liberal party. In the 2019 federal election, the voting rate of the Liberal party is 33.1% (Stephenson et al., 2021) which is very close to our estimation. Thus, we have said that the result of our model is reasonable.

However, the other parties prediction seem not that accurate which reflect our models' limitation. We saw there are some unexpected values far away from the data in the previous election result. In our search for these reasons, we found that this situation was caused by two limitations.

For the Bloc Québécois party, the result shows that there is only a 0.012912% voting rate which is far away from the voting rate in the 2021 election of 7.6%. the reason that causes this situation is that in our model, we delete some options in the question to make our survey data fit with our census data. For example, we remove the variable "speak both French and English in home language variable in survey data due to it has small proportion. Yet, it maybe influential for the model of some parties. More specifically, it might cause some people who voted for Bloc Québécois were removed from the data set. Also, for the NDP, green party, and people party. The reason for the difference between the estimate and the previous year's data is that there do not have many questions that can favor the people who want to vote for them. Such as the green party stands for protecting the environment. But in the survey, there is no option for people who care about the environmental issue. Thus, it causes this survey cannot capture those voting citizens. However, despite these limitations, we can improve it in our future study by conducting a new survey that can fit the census.

Conclusions

Throughout all these parts of the research, we first introduce our research question that which party is most likely to win in the 2025 federal election. based on the 2019 Canadian election survey and the 2017 General Social Survey. We use to build the logistic regression models for each party by using 7 predictors. Then, based on these models, we use post-stratification to get an estimate of the voting rate for each party. In our result, we can see that the liberal party is most likely to win in the next federal election.

In the big picture, we can say that after getting this result, we can expect that Canada will continue their policy since the liberal party will still hold the power in the country. And Canada is also expected to keep their economic growth.

As for the weakness and limitations in our research, we found that the census data and survey data are from different years. One is collected in 2019 and one is collected in 2017. This difference may cause some people to change their answers to some questions. Also, there are some questions in the census do not have in the survey. Which may cause the miss answering. Thus, in the future, we will conduct more accurate research on predicting federal election results by construction and using the survey based on the census. Then, we can get a more accurate sample use for our logistic regression.

Bibliography

- Allaire, J.J., et. el. *References: Introduction to R Markdown*. RStudio. <https://rmarkdown.rstudio.com/docs/>. (Last Accessed: January 15, 2021)
- Armstrong, D. A., Lucas, J., & Taylor, Z. (2022). The Urban-Rural Divide in Canadian Federal Elections, 1896–2019. *Canadian Journal of Political Science*, 55(1), 84–106. <https://doi.org/10.1017/S0008423921000792>
- Dekking, F. M., et al. (2005) *A Modern Introduction to Probability and Statistics: Understanding why and how*. Springer Science & Business Media.
- Grolemund, G. (2014, July 16) *Introduction to R Markdown*. RStudio. https://rmarkdown.rstudio.com/articles_intro.html. (Last Accessed: January 15, 2021)
- Heard, A. (n.d.). Canadian election results by party 1867 to 2021. Canadian Election Results: 1867-2021. Retrieved from <https://www.sfu.ca/~aheard/elections/1867-present.html>
- Ma, C. (n.d.). Voting behaviour in Canada. The Canadian Encyclopedia. Retrieved November 30, 2022, from <https://www.thecanadianencyclopedia.ca/en/article/electoral-behaviour>
- Stephenson, L. B., Harell, A., Rubenson, D., & Loewen, P. J. (2021). Measuring Preferences and Behaviours in the 2019 Canadian Election Study. *Canadian Journal of Political Science*, 54(1), 118–124. <https://doi.org/10.1017/S0008423920001006>
- Uppal, S., & LaRochelle-Côté, S. (2015, November 27). Factors associated with voting. Statistics Canada: Canada’s national statistical agency / Statistique Canada: Organisme statistique national du Canada. Retrieved November 30, 2022, from <https://www150.statcan.gc.ca/n1/pub/75-001-x/2012001/article/11629-eng.htm>
- Welch, S., & Hibbing, J. (1992). Financial Conditions, Gender, and Voting in American National Elections. *The Journal of Politics*, 54(1), 197–213. <https://doi.org/10.2307/2131650>

Appendix

```
glimpse(census_data_final)
```

```
## Rows: 19,229
## Columns: 7
## $ age                <dbl> 55, 53, 66, 82, 30, 65, 61, 82, 66, 27, 18, 42~
## $ gender             <chr> "Female", "Male", "Female", "Female", "Male", ~
## $ family_income      <chr> "medium income", "medium income", "medium inco~
## $ current_living_province <chr> "Quebec", "Manitoba", "Ontario", "Alberta", "Q~
## $ education_level    <chr> "Less than high school and high school", "Coll~
## $ home_language      <chr> "French only", "English only", "French only", ~
## $ life_feeling       <chr> "Very satisfied", "Very satisfied", "Very sati~
```

```
glimpse(survey_data_final)
```

```
## Rows: 605
## Columns: 13
## $ vote_Liberal       <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0~
## $ vote_Conservative  <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1~
## $ vote_NDP           <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0~
## $ vote_Bloc_QB       <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ vote_Bloc_Green    <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0~
## $ vote_Bloc_People   <dbl> 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ age                <dbl> 41, 20, 29, 29, 52, 70, 39, 73, 34, 45, 40, 35~
## $ gender             <chr> "Male", "Male", "Female", "Male", "Male", "Mal~
## $ family_income      <chr> "medium income", "medium income", "medium inco~
## $ current_living_province <chr> "Quebec", "Quebec", "Quebec", "Quebec", "Quebe~
## $ education_level    <chr> "Master and PHD degree", "Less than high schoo~
## $ home_language      <chr> "French only", "French only", "French only", "~
## $ life_feeling       <chr> "Somehow satisfied", "Somehow satisfied", "Som~
```

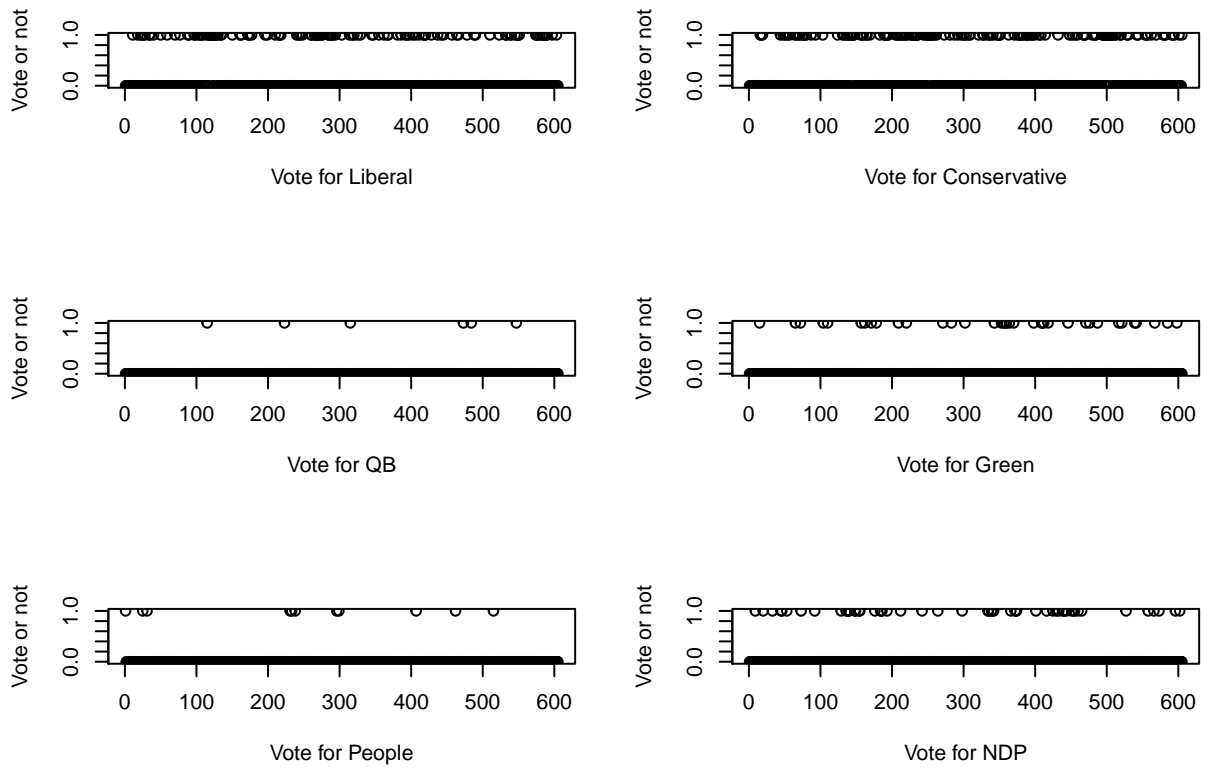


Fig.5

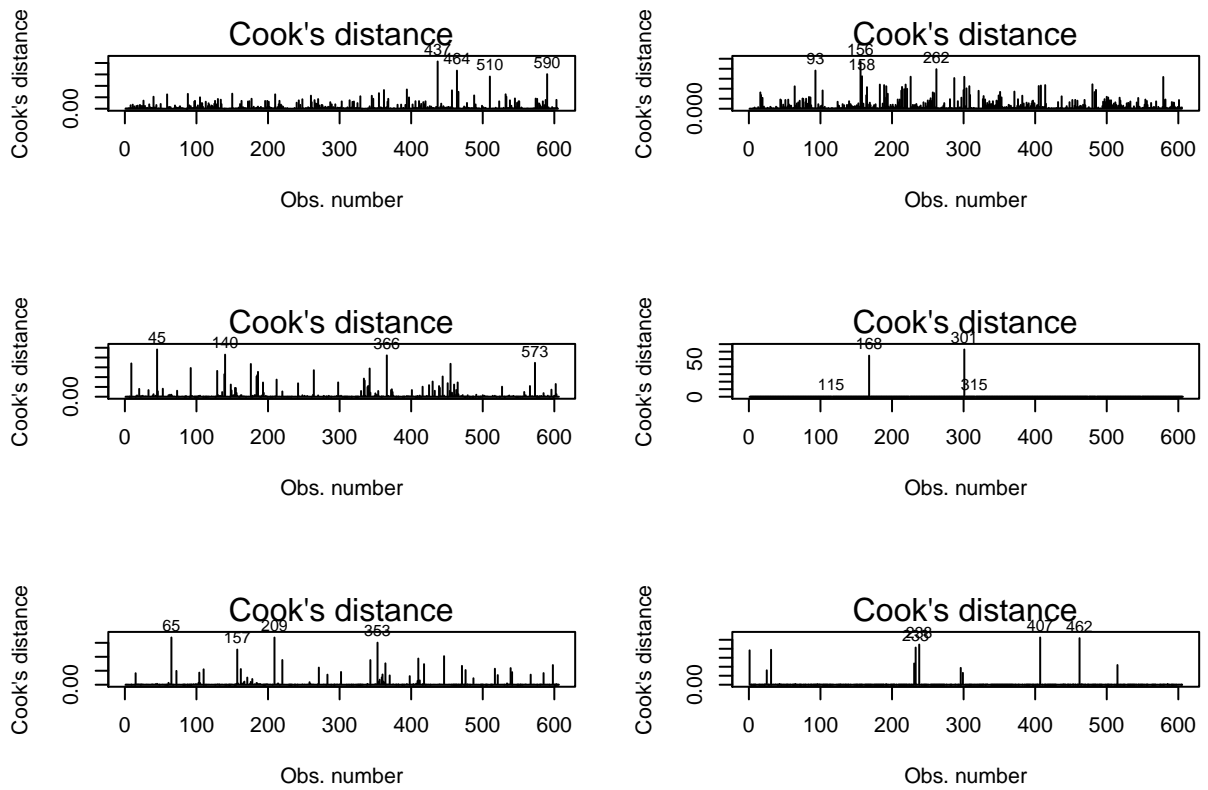


Fig.6