Identification of How the 2019 Canadian Federal Election Would Have Been Different If 'Everyone' Had Voted Using Propensity Score Matching, Multinomial Logistic Regression and Post-stratification

Xiaoxuan Han 1005868161

12/22/2020

Code and data supporting this analysis is available at: https://github.com/rosehan2001/STA304FinalProject.git

Abstract

In 2019 Canadian Federal Election, liberal party has won among all the other parties and Justin Trudeau was able to take another opportunity to lead Canada. However, it has come to our attention that the election result might be different if "everyone" has voted since some people might fail to vote due to unexpected circumstances on election day. In this scenario, "everyone" will be considered as all the people who are above 18 years old and answered General Social Survey in 2017. In this paper of study, we believe the turnout on election day might play an important role in affecting the final election result, so we will show how to properly predict the election result by first performing propensity score mathcing and creating multinomial logistic regression models based on 2019 Canadain Election Study Online Survey Data, and perform post-stratification based on 2017 General Social Survey Family Data to obtain estimates of popular votes for each political party. By comparing the esitmated prediction result and the real popular votes, we conclude by stating that the turnout based on models and result is significant in a way that different turnout will possibly lead to different election results.

Keywords

2019 Canadian Federal Election, Propensity Score Matching, Casual Link Observational Study, Multinomial Logistic Regression, Post-Stratification, Turnout

Introduction

During election season, prediction of election result based on various resources and surveys have been useful in analyzing the political trends over years. It is crucial to know that the actual eletion result might not be reprentative enough for the entire population since not all Canadian people have successfully voted on election day due to various reasons. Thus, we want to explore if the election result is going to be different if "everyone" has voted on election day. In order to achive this goal, we have found large-scale surveys Canadian Election Survey (CES), which have been conduted during each Canadian federal election since 1965. In these surveys, valuable information including voters' intention, voters' personal information, the major parties that voters' leaning towards and etc. are available. The subset of 2019 online survey data have been chosen as the survey data set in this paper of study. In order to obtain accurate prediction result of popular vote from everyone, there are few key points that need to be noticed. First, some casual inference of observational data need to be incorporated in the analysis. Second, identifying what variables might be relevent in creating prediction models is helpful in getting more accurate response. Third, choosing the appropriate model for prediction is significant since different statistical models usually lead to different results.

Over the years, there tends to be a noticeable gap between young people and old people, from both mentally and physically. Therefore, it is common for them to hold different opinions towards the world, which also includes their voting choice. This phenomenon indicates a potentail casual inference, and a popular way for making casual inference is propensity score matching. Propensity score matching allows us analyze observational data and mimicing particular characteristcs of a randomized controlled cell, where we have the oppourtunity to descrive casual inference between treatment and outcome. This method will be used on finding out if there is a casual link between whether the indiviual is young and old and whether this person votes for a particualr party. Important explanatory variables which highlight an individual's personal information including age(in two groups), sex, education and province will be selected for the prediction model. Multinomial logistic regression models will be selected as our prediction model since we are predicting a binary response based on two or more independent variables, where the log odds of the outcome is modeled as a linear combination of the predictor variables. By using this type of model which paired with propensity score mathcing, we will perform post-stratification, which is the final step of predicting popular vote for each political party. Post-stratification is important in generating more accurate result is that it refers to the process of adjusting the estimates, essentially a weighted average of estimates from all possible combinations of attributes. Eventually, we did find out a difference between estimates of prediction election result and actual election result, so we performed further analysis regarding how voters' turnout based on model and result is important in election result.

In this paper, 2019 Canadian Election Study (CES) Online Survey Data and 2017 General Social Survey (GSS) Family Data will be used to investigate how propensity score matching could be used to make inference on the casual inference that can be made between age and voting result, find out what prediction result we will get for each party by creating multinomial logistic regression models, and perform post-statification on census data (GSS). In the Methodology section, key points on the two data sets will be introduced, as well as the models used to perform propensity score matching analysis and multinomial logistic regression analysis. Models results and estimated prediction of popular vote for each party will be provided in the Results section. Summary, conclusions, weaknesses and what we can do in the future regarding the turnout of models and result will be presented in the Discussion section. All the resources and materials that were referenced will be presented both in lines and in the Reference Section. Some summart tables of model results have been added the Appendix Section.

Methodology

Data

2019 Canadian Election Study Online Survey Data has been used as the survey data for performing propensity score matching and building prediction models for each political party. It has 37822 observations in total and after data cleaing process, 36363 observations are left for usage. The original data set includes 620 variables in total, but in order to make prediction models, the data size will be reduced and only relevant variables will be selected. We decide to focus on the respondent's personal characteristics and experience, so we create the summary of the respondent's age, the gender of the respondent, the highest education level completed, province in which the repsondent lives. It can be seen that the overall proportion of older people tends to be smaller than the proportion of younger people. There is also greater proportion of people who completed technical, community college, CEGEP, College Classique and who achieve bachelor's degree. There is a significant proportion amount of respondents who live in Ontaio. Also, there tends to be more female than male, where repsondents of other genders only occupies a smaller proportion.

Figure 1. Baseline Characteristics of 2019 Canadian Election Study Online Survey Data After Data Cleaning

```
##                                                                   Stratified by age
##                                                                    0
##   n                                                                17903
```

```
##   sex = Male (%)                                                 6867 (38.4)
##   province (%)
##      Alberta                                                     2245 (12.5)
##      British Columbia                                            1858 (10.4)
##      Manitoba                                                     783 ( 4.4)
##      New Brunswick                                                406 ( 2.3)
##      Newfoundland and Labrador                                    313 ( 1.7)
##      Nova Scotia                                                  456 ( 2.5)
##      Nunavut                                                       15 ( 0.1)
##      Ontario                                                     7328 (40.9)
##      Prince Edward Island                                          67 ( 0.4)
##      Quebec                                                      3775 (21.1)
##      Saskatchewan                                                 657 ( 3.7)
##   education (%)
##      Bachelor's degree (e.g. B.A., B.Sc., LL.B.)                 4791 (26.8)
##      College, CEGEP or other non-university certificate or trade   5827 (32.5)
##      High school diploma or a high school equivalency certificate  2528 (14.1)
##      Less than high school diploma or its equivalent               858 ( 4.8)
##      University certificate or diploma below the bachelor's level  1688 ( 9.4)
##      University certificate, diploma or degree                    2211 (12.3)
##                                                                  Stratified by age
##                                                                   1
##   n                                                               8460
##   sex = Male (%)                                                  4541 (53.7)
##   province (%)
##      Alberta                                                       939 (11.1)
##      British Columbia                                             1144 (13.5)
##      Manitoba                                                      385 ( 4.6)
##      New Brunswick                                                 175 ( 2.1)
##      Newfoundland and Labrador                                     117 ( 1.4)
##      Nova Scotia                                                   249 ( 2.9)
##      Nunavut                                                         4 ( 0.0)
##      Ontario                                                      3162 (37.4)
##      Prince Edward Island                                           35 ( 0.4)
##      Quebec                                                       1943 (23.0)
##      Saskatchewan                                                  307 ( 3.6)
##   education (%)
##      Bachelor's degree (e.g. B.A., B.Sc., LL.B.)                  1814 (21.4)
##      College, CEGEP or other non-university certificate or trade  2707 (32.0)
##      High school diploma or a high school equivalency certificate 1456 (17.2)
##      Less than high school diploma or its equivalent               415 ( 4.9)
##      University certificate or diploma below the bachelor's level  932 (11.0)
##      University certificate, diploma or degree                    1136 (13.4)
##                                                                  Stratified by age
##                                                                   p      test
##   n
##   sex = Male (%)                                                  <0.001
##   province (%)                                                    <0.001
##      Alberta
##      British Columbia
##      Manitoba
##      New Brunswick
##      Newfoundland and Labrador
##      Nova Scotia
```

```
##       Nunavut
##       Ontario
##       Prince Edward Island
##       Quebec
##       Saskatchewan
##    education (%)                                                      <0.001
##       Bachelor's degree (e.g. B.A., B.Sc., LL.B.)
##       College, CEGEP or other non-university certificate or trade
##       High school diploma or a high school equivalency certificate
##       Less than high school diploma or its equivalent
##       University certificate or diploma below the bachelor's level
##       University certificate, diploma or degree
```

We cleaned 2019 Canadian Election Study Online Survey Data by selecting four relevant categorical variables for model prediction, which are the respondent's age, the gender of the respondent, the highest education level completed, province in which the repsondent lives. All missing values have been removed from the original data set since missing values will make model prediction hard and complicated. Since an obvious difference in age has been noticed from original dataset, we stratify age into two groups to make a baseline charactristics tableand and check if these groups are similar at baseline. As it turns out from the table, two groups indeed show similar proportions in sex, province, and education, so we will be more confident that observed differences in outcomes between the groups have to do with the intervention rather than confounding factors.

The census data that we choose to perform post-stratification on is 2017 General Social Survey Family Data. This data set was obtained from Statistics Canada, and we performed data cleaning process to reduce its size and manipulate its variables to make it easier for post-stratification process. Its target population is all non-institutionalized persons 15 years of age or older, living in the 10 provinces of Canada. Since the eligible age for voting during elections is 18, all the respondents who are below 18 have been removed from the data set. Also, since this data was collected with voluntary response and the overall response rate is 52.4%, all missing values have been removed from the data.

Model

We will be using multinomial logistic regression models paried with propensity score mathcing to model the proportion of voters who will vote for each party. The reason we chose this multinomial logistic regression model is that we are predicting a binary response based on two or more independent variables, where the log odds of the outcome is modeled as a linear combination of the predictor variables. We will be using age (in two groups), sex, education, and province which are all recorded as categorical variables, to perform the modeling, and age will be the propensity. We will run the seven multinomial logistic regression model for each political party by using the function `glm()` and matched data generated from propensity score matching, and then save the descriptive statistics into a summary table.

We will first build a logistic regression model for the treatment, which is age (in two groups), to get the propensity score. The multinomial logistic regression model for logistic regression model for the treatment is:

$$y_{age} = \beta_{intercept} + \beta sex x_{sex} + \beta_{province} x_{province} + \beta_{education} x_{education} + \epsilon$$

. y_{age} will the propensity score, and x_{sex}, $x_{province}$, $x_{education}$ are all predictor variables of sex, province, and education repectively. They correspond to estimates of βsex, $\beta_{province}$, $\beta_{education}$ respectively in this model. Propensity score matching aims to create groups of treated and control subjects which have similar covariate values so that subsequent comparisons, made within these two matched groups, are not confounded by differences in covariate distributions. We estimate gaps for old and young people and analyze average treatment effects on the outcome variable, which is vote choice. In order to obtain matched pairs of treated and untreated subjects when matching with propensity score, we first choose between matching without replacement. When an untreated subject has been matched to a treated subject, a matched pair has been created. This process repeats for many times, until we get a full propensity matched data set. Based on the propensity score matched data, we will then be able to build multinomial logistic regression models to predict votechoice for each party using explanatory variables age, sex, province and educaton.

The multinomial logistic regression model for predictin the proportion of voters who will vote forLiberal Party is:

$$y_{liberalparty} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_{17} x_{17} + \epsilon$$

. $p_{liberalparty}$ represents the proportion of voters who will vote for Liberal Party. y_1 represents the response variable in the multinomial logistic regression. It represents the log odds of the proportion of voters who will vote for Liberal Party. Similarly, β_0 represents the intercept of the model, which is the fixed baseline intercept representing the proportion of voters who are younger than 60, who are female, who lives in Alberta, who completed Bachelor's degreem, and will vote for liberal party. β_1 is the estimate of the age group categories "below 60 years old", and "above is 60 years old", respectively, with "below 60 years old" as the reference category; β_2 is the estimate of sex,with "Female" as the reference category; β_3 to β_{12} are the estimates of province categories, with "Alberta" as the reference category; β_{13} to β_{17} are estimates of education categories, with "Bachelor's Degree" as the reference category. The predictor variables are age (in two groups), sex, education and province.

Another six multinomial logistic regression models for predicting the proportion of voters who will for Conservative Party, NDP Party, Bloc Quebecois Party, Green Party, People's Party, and Another Party have also been built by using the same logic as building the model for Liberal Party. There models results are included in the Appendix. The multinomial logistic regression model for predicting the proportion of voters who will vote for Conservative Party is:

$$y_{conservative} = \beta_{18} + \beta_{19} x_{19} + \beta_{20} x_{20} + ... + \beta_{35} x_{35} + \epsilon$$

. The multinomial logistic regression model for predicting the proportion of voters who will vote for NDP Party is:

$$y_{ndp} = \beta_{36} + \beta_{37} x_{37} + \beta_{38} x_{38} + ... + \beta_{53} x_{53} + \epsilon$$

. The multinomial logistic regression model for predicting the proportion of voters who will vote for Bloc Quebecois Party is:

$$y_{blocquebecois} = \beta_{54} + \beta_{55} x_{55} + \beta_{56} x_{56} + ... + \beta_{71} x_{71} + \epsilon$$

. The multinomial logistic regression model for predicting the proportion of voters who will vote for Green Party is:

$$y_{green} = \beta_{72} + \beta_{73} x_{73} + \beta_{74} x_{74} + ... + \beta_{89} x_{89} + \epsilon$$

. The multinomial logistic regression model for prediting the proportion of voters who will vote for People's Party is:

$$y_{people's} = \beta_{90} + \beta_{91} x_{91} + \beta_{92} x_{92} + ... + \beta_{107} x_{107} + \epsilon$$

. The multinomial logistic regression model for predicting the proportion of voters who will vote for Another Party is:

$$y_{anotherparty} = \beta_{108} + \beta_{109} x_{109} + \beta_{110} x_{110} + ... + \beta_{125} x_{125} + \epsilon$$

. Each p_i represents the proportion of voters who will vote for "i" political party, and y_i is the response variable in each multinomial logistic regression, where i = Conservative Party, NDP Party, Bloc Quebecois Party, Green Party, People's Party,Another Party. The interpretation of parameters from β_{19} to β_{125} will follow the same logic as the interpreation for parameters from β_0 to β_{17}. The predictor variables are age (in two groups), sex, education and province, which are exactly the same in each model.

Post-strafication Predication

In order to estimate the proportion of people who will vote for each political party, we need to perform a post-stratification analysis. Post-stratification analysis is performed with the aim of conducting inference on a large data sample that may not necessarily be representative of the target population. In order to conduct post-stratification analysis, we need to collect large amounts of demographic data along with the sample. We partition the data by cells with all the possible combinations of demographic characteristics, estimate a response for each cell, and estimate the response of the target population by weighting each cell by the number of data points in each cell. Here we create cells based on the chosen demographic characteristics, which are age (in two groups), sex, province and education. The data is partitioned into different cells, which have at least one person with the described characteristics. The weighting of each cell is calculated by multiplying the response estimate of each cell by the respective population size of that cell, then adding together all of these values, and finally dividing the sum by the entire population size.

The formula for the post-stratification calculation is as follows:

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

, where $\hat{y}_j$ is the estimated response for the j^{th} cell, N_j is the number of samples per cell, and $\hat{y}^{PS}$ is the estimated response for the target population.

Results

Model Results

Figure 2: Descriptive statistics of the model of Liberal Party regression estimates

term	estimate	std.error	statistic	p.value
(Intercept)	-1.3124187	0.0734368	-17.871401	0.0000000
age	0.1620719	0.0331554	4.888254	0.0000010
sexMale	-0.1992773	0.0334903	-5.950290	0.0000000
provinceBritish Columbia	0.7983201	0.0767055	10.407605	0.0000000
provinceManitoba	0.7502775	0.1001603	7.490766	0.0000000
provinceNew Brunswick	1.2567443	0.1261512	9.962210	0.0000000
provinceNewfoundland and Labrador	1.7705604	0.1460856	12.120017	0.0000000
provinceNova Scotia	1.5639597	0.1094030	14.295396	0.0000000
provinceNunavut	2.5809651	0.8418870	3.065691	0.0021717
provinceOntario	1.1512076	0.0670708	17.164077	0.0000000
provincePrince Edward Island	1.3516413	0.2531679	5.338913	0.0000001
provinceQuebec	1.0389904	0.0705685	14.723151	0.0000000
provinceSaskatchewan	-0.1313817	0.1284561	-1.022775	0.3064141
educationCollege, CEGEP or other non-university certificate or trade	-0.3959932	0.0457616	-8.653401	0.0000000
educationHigh school diploma or a high school equivalency certificate	-0.5678341	0.0546020	-10.399519	0.0000000
educationLess than high school diploma or its equivalent	-0.5375360	0.0854540	-6.290354	0.0000000
educationUniversity certificate or diploma below the bachelor's level	-0.1277642	0.0600268	-2.128454	0.0332995
educationUniversity certificate, diploma or degree	0.0896798	0.0555033	1.615756	0.1061471

Only the summry table of estimates of model estimating the proportion of voters voting for Liberal Party is included in Figure 2, all the other summary results have all been added to Appendix. In each of the six prediction models, the interpretation of each estimate will follow the same principles. For example, with significance of variables in this analysis determined by the p-value threshold of $p < 0.05$, it can be seen in Figure 1 that province British Columbia is significant and has an estimate of $\beta_3 = 0.7983201$. The regression estimates are the log odds of the voting estimate, therefore the interpretation of this result is that the likelihood of voting for Liberal Party of respondents who live in British Columbia is $\exp(\beta) = 2.2218053$ times that of voting for Liberal Party of people living in Alberta, when controlling for all other variables.

For the model estimating the proportion of voters voting for Liberal Party, all the categories are significant based on p-values except for categories province:Saskatchewan and education: University certificate, diploma or degree; for the model estimating the proportion of voters voting for Conservative Party, all the categories are significant based on p-values except for category education: University certificate or diploma below the bachelor's level; for the model estimating the proportion of voters voting for NDP Party, all the categories are significant based on p-values except for categories province:Nunavut, province:Prince Edward Island, province:Quebec, education:College, CEGEP or other non-university certificate or trade, education:University certificate, diploma or degree; for the model estimating the proportion of voters voting for Green Party, all the categories are significant based on p-values except for categories province:Newfoundland and Labrador, province:Saskatchewan, and all the education levels; for the model estimating the proportion of voters voting for People's Party, only age, sex, province New Brunswick, educationCollege, CEGEP or other non-university certificate or trade, educationHigh school diploma or a high school equivalency certificate, education:Less than high school diploma or its equivalent are significant categories; for the model estimating the proportion of voters voting for Another Party, only province:Saskatchewan is the significant category. It is noticeable that in the model estimating the proportion of voters voting for Bloc Quebecois Party, all the p-values are extremely large, so this model is bad at predicting popular vote and is very likely to be an invalid model. The predicting ability of a model is weaker with fewer significant categories, so models that have many insignificant catergories need to be improved and we shall explain this in Discussion section.

Post-stratification Results

Figure 3: Table of Estimated Popular Vote

	Popular Vote
Liberal Party	0.2827609
Conservative Party	0.3672402
NDP Party	0.1834420
Bloc Quebecois Party	0.0212873
Green Party	0.1182828
People's Party	0.0285573
Another Party	0.0075643

We estimate that the proportion of voters in favour of voting for Liberal Party to be 0.283, for Conservative Party is 0.367, for NDP Party is 0.183, for Bloc Quebecois Party is 0.021, for Green Party is 0.118, for People's Party is 0.029, for Another Party is 0.008. These estimates are based off our post-stratification analysis of the proportion of voters in favour of each party modelled by a multinomial logistic regression model respectively, which accounted for age (in two groups), sex, province and education. Figure 3 shows the estimated popular vote for each political party.

Discussion

Summary

In this paper of study, both 2019 CES Online Survey Data and 2017 General Socail Survey Data have been cleaned before conducting any statitsical analysis. We have focused on performing multinomial logistic regression paired with propensity score matching to estimate the predicted popular vote for each political party in post-stratification process. During propensity score matching, we have examined the casual link between whether the individual is young and whether the individual votes for a particular party. After matching similar subjects, we obtained the matched data which has been reduced from the cleaned version of

2019 CES Online Survey Data, and used it to build our multinomial logistic regression models. Whether or not voting for a particular party has been considered as a binary response variable for each model, and age(in two groups), sex, province, and education are all categorical explanatory variables which are used for predicting the response vote choice. Each model has been analyzed regarding whether or not its coefficients are significant and are useful for predicting results. We eventually performed post-stratification process on 2017 General Social Surey Data to get the desired estimates of popular vote for each political party.

Conclusions

Figure 4: Popular Votes of Actual Result vs. Prediction result

	Actual Popular Vote	Predicted Popular Vote
Liberal Party	0.3312	0.2827609
Conservative Party	0.3434	0.3672402
NDP Party	0.1598	0.1834420
Bloc Quebecois Party	0.0763	0.0212873
Green Party	0.0655	0.1182828
People's Party	0.0164	0.0285573
Another Party	0.0067	0.0075643

In conclusion, the results show that if "everyone" (as how we defined in the Introduction section) has voted, Conservative Party will gain more popular vote and win the 2019 Canadian Federal Election. However, in reality, Liberal Party won the federal election in 2019. During the study, by first performing propensity score matching, where age (in two groups) is set as propensity, and obtained a propensity score matched data set, and then building multinomial logistic regression models for each Party respectively, and eventually conducting post-stratification on census data, there is sufficient evidence to show that the methods we employed during the entire study is convincing in supporting this reuslt.

The actual election result is 0.3312 for Liberal Party, 0.3434 for Conservative Party, 0.1598 for NDP Party, 0.0763 for Bloc Quebecois Party, 0.0655 for Green Party, 0.0164 for People's Party, and 0.0067 for Another Party (Wikipedia contributors, 2020). In Figure 4, we make comparison between prediction estimates of popular vote and the actual popular vote. More popular vote for Conservative Party compared to Liberal Party reveals a potential fact that a greater amount of people who did not vote on election day favour Conservative Party. For the estimated popular vote for other political parties except for Bloc Quebecois Party, the prediction and actual election resulst are similar. The huge difference between the prediction result and actual result for Bloc Quebecois Party is due to its inappropriate prediction model. Also, we see that the turnout on election day is important for the election results since different turnouts might cause different final election results. A low turnout of a certain group will cause an disproportinate affect on the popular vote. In 2019 Canadian Federal Election, we can deduce that people who support Conservative Party has a low turnout on election day based on the model result. It is true that there exist various factors which can affect the turnout of a voter, such as timing of the election day, weather and etc., so the election result should be subjected to change due to different turnout rate of voters.

Weakness & future steps

Weaknesses in this study are still need to discussed. First, an inappropriate age boundary might have been set to divide people into young and old groups. When perfroming propensity score matching, people have been assgined to two age groups, where 60 years old is a boundary. It defines people who are below 60 years old as young, and who are above 60 years old as old. However, the amount of people who are below 60 years old and the amount of people above 60 years old are not the same in this data set, so this causes potential errors due to the unequivalent amount of units in treatment and control groups. Also, the boundary of age 60 might contain bias since people's mental age is sometimes different from their physical age, so it will be better if we could find out a more fair way to split people into two age groups. Second, the census data that was used for post-stratification process is from 2017 rather than 2019. This might create potential change of an individual. For example, a person who did not finish Bachelor's degree might finish it in 2019. At this time, our prediction result becomes inaccurate if we still use census data from 2017 since the value of explanatory variable of this person has changed in 2019. Third, when cleaning CES Online Survey Data, we omitted sex category "Other". The reason behind doing this was that there are only two categories of sex in 2017 General Social Survey Data. However, this category does play a role for prediction in a model, and it might cause a substantial change in the prediction result. For example, when a party strongly favours the policy of LGBTQ, then people who choose sex as "Other" might vote for this party with a greater possibility. Last, only four explanatory variables have been chosen due to the limited amount of time. While these partitioned our data into cells with distinguishable demographic characteristics, in real life, there are many more factors (i.e. income, employment status, etc.) and variables that we need to consider when predicting the voting outcome. In terms of our data cleaning process, we took out the "respondent skipped", or the NA values from the original data set. This means that our analysis does not account for the opinions of those who did not choose to reveal their personal information, and it may cause the non-sampling error.

More importantly, there are few things which we can do in the future to improve this statistcal analysis. When comparing our predicted result with the actual election result, it can be seen that the predicted results and actual results are different. This finding supports what we have been exploring and expecting in the study, but we still need to check if ant discrepency exists while predicting popular vote. One way to make our study more practical would be to conduct post-hoc analysis and compare the results from the actual election outcome. Follow-up surveys or exit polls can be conducted to either confirm our analysis or provide insight for ways to improve our analytical methods. We could use a least squares regression model to contrast and compare the actual observed data and our predicted data. The surveys and exit polls could contain questions such as "Did you vote on election day?" "If so, is there a specific reason that makes you fail to vote?" or "Are you satisfied with the election result?". The answers to these question will let us know more about a voter's personal opinion on election. For example, we could try to find out if there is a casual link between whether people vote on election day or not and whether they are satisfied with the result, and which party they think they will vote based on current opinion of election result. By knowing more information regarding how people's thoughts have changed during before and after the election, we will be able to perform more strict model predictions, such as a multilevel regression post-stratification (MRP) model, which requires fitting two nested multilevel logistic regressions for estimating candidate support in each cell, and then weighing cell-level estimates by the proportion of the electorate in each cell and aggregated to the appropriate level. Lastly, another way to predict more accurate election result can be predicting the turnout based on previous years election study, and then use the voters who are predicted to turn out on election day to perform predictions.

References

1. Austin, Peter C. "An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies." Multivariate Behavioral Research, Taylor & Francis, May 2011, www.ncbi.nlm.nih.gov/pmc/articles/PMC3144483/.

2. "How to Create Table 1 Summary Information." Exploratory, exploratory.io/note/kanaugust/How-to-Create-Table-1-Summary-Information-YYf5BPf4Xs.

3. Government of Canada, Statistics Canada. "General Social Survey – Family (GSS)." Government of Canada, Statistics Canada, 13 Nov. 2020, www.statcan.gc.ca/eng/survey/household/4501.

4. Masters, Anthony B. "MRP Estimates and the 2019 General Election." Medium, Medium, 27 Nov. 2019, anthonybmasters.medium.com/mrp-estimates-and-the-2019-general-election-9ac1794120d6.

Appendix

Figure A1: Descriptive Statistics of the Model of Conservative Party Regression Estimates

term	estimate	std.error	statistic	p.value
(Intercept)	0.2267824	0.0648287	3.498181	0.0004684
age	0.1753509	0.0340196	5.154401	0.0000003
sexMale	0.3964565	0.0345817	11.464338	0.0000000
provinceBritish Columbia	-1.3819046	0.0668004	-20.687079	0.0000000
provinceManitoba	-0.9297431	0.0881236	-10.550447	0.0000000
provinceNew Brunswick	-1.4746832	0.1274750	-11.568410	0.0000000
provinceNewfoundland and Labrador	-1.7491704	0.1587251	-11.020124	0.0000000
provinceNova Scotia	-1.9474012	0.1198205	-16.252655	0.0000000
provinceNunavut	-2.5401858	1.0829185	-2.345685	0.0189921
provinceOntario	-1.3205487	0.0562077	-23.494106	0.0000000
provincePrince Edward Island	-2.0763309	0.3058777	-6.788109	0.0000000
provinceQuebec	-2.1553759	0.0644344	-33.450692	0.0000000
provinceSaskatchewan	-0.3727517	0.0960627	-3.880296	0.0001043
educationCollege, CEGEP or other non-university certificate or trade	0.3286918	0.0475118	6.918114	0.0000000
educationHigh school diploma or a high school equivalency certificate	0.4064257	0.0550837	7.378325	0.0000000
educationLess than high school diploma or its equivalent	0.2350686	0.0844402	2.783848	0.0053718
educationUniversity certificate or diploma below the bachelor's level	-0.0734560	0.0652093	-1.126464	0.2599690
educationUniversity certificate, diploma or degree	-0.1753314	0.0612373	-2.863149	0.0041945

Figure A2: Descriptive Statistics of the Model of NDP Party Regression Estimates

term	estimate	std.error	statistic	p.value
(Intercept)	-1.9323555	0.0977871	-19.7608347	0.0000000
age	-0.6428406	0.0480955	-13.3659261	0.0000000
sexMale	-0.3564699	0.0471880	-7.5542557	0.0000000
provinceBritish Columbia	0.7244262	0.0966937	7.4919677	0.0000000
provinceManitoba	0.4750626	0.1293415	3.6729335	0.0002398
provinceNew Brunswick	-0.4427943	0.2350940	-1.8834775	0.0596357
provinceNewfoundland and Labrador	0.5736647	0.1974895	2.9047853	0.0036751
provinceNova Scotia	0.3907941	0.1535696	2.5447356	0.0109361
provinceNunavut	-10.3024225	121.1658621	-0.0850274	0.9322396
provinceOntario	0.4526365	0.0872123	5.1900551	0.0000002

term	estimate	std.error	statistic	p.value
provincePrince Edward Island	-0.1346751	0.4368824	-0.3082641	0.7578814
provinceQuebec	-0.1429117	0.0986624	-1.4484911	0.1474798
provinceSaskatchewan	0.7769684	0.1310745	5.9276871	0.0000000
educationCollege, CEGEP or other non-university certificate or trade	0.1253695	0.0673246	1.8621639	0.0625800
educationHigh school diploma or a high school equivalency certificate	0.1691796	0.0775553	2.1814067	0.0291533
educationLess than high school diploma or its equivalent	0.4063093	0.1101964	3.6871378	0.0002268
educationUniversity certificate or diploma below the bachelor's level	0.2640463	0.0864052	3.0559062	0.0022438
educationUniversity certificate, diploma or degree	0.0940110	0.0841669	1.1169587	0.2640120

Figure A3: Descriptive Statistics of the Model of Bloc Quebecois Party Regression Estimates

term	estimate	std.error	statistic	p.value
(Intercept)	-22.1157184	6.676287e+02	-0.0331258	0.9735743
age	0.7828700	7.595140e-02	10.3075109	0.0000000
sexMale	0.1315377	7.656800e-02	1.7179206	0.0858111
provinceBritish Columbia	0.0042236	9.014343e+02	0.0000047	0.9999963
provinceManitoba	-0.0084198	1.232918e+03	-0.0000068	0.9999946
provinceNew Brunswick	-0.0127977	1.694586e+03	-0.0000076	0.9999940
provinceNewfoundland and Labrador	0.0303672	2.005774e+03	0.0000151	0.9999879
provinceNova Scotia	0.0238276	1.454987e+03	0.0000164	0.9999869
provinceNunavut	-0.1092106	1.094727e+04	-0.0000100	0.9999920
provinceOntario	0.0103046	7.602198e+02	0.0000136	0.9999892
provincePrince Edward Island	-0.0101240	3.542750e+03	-0.0000029	0.9999977
provinceQuebec	20.5042865	6.676287e+02	0.0307121	0.9754991
provinceSaskatchewan	-0.0112575	1.342280e+03	-0.0000084	0.9999933
educationCollege, CEGEP or other non-university certificate or trade	-0.0690741	1.105009e-01	-0.6251000	0.5319054
educationHigh school diploma or a high school equivalency certificate	0.2545152	1.118973e-01	2.2745418	0.0229334
educationLess than high school diploma or its equivalent	0.2199586	1.823551e-01	1.2062104	0.2277364
educationUniversity certificate or diploma below the bachelor's level	0.1334896	1.311068e-01	1.0181749	0.3085948
educationUniversity certificate, diploma or degree	-0.1630302	1.413540e-01	-1.1533469	0.2487680

Figure A4: Descriptive Statistics of the Model of Green Party Regression Estimates

term	estimate	std.error	statistic	p.value
(Intercept)	-2.7450423	0.1272515	-21.5717836	0.0000000
age	-0.2923264	0.0551828	-5.2974227	0.0000001
sexMale	-0.2429000	0.0552153	-4.3991446	0.0000109
provinceBritish Columbia	1.2268606	0.1240151	9.8928317	0.0000000
provinceManitoba	0.7254578	0.1643168	4.4149959	0.0000101
provinceNew Brunswick	1.3993777	0.1808000	7.7399227	0.0000000
provinceNewfoundland and Labrador	-0.4956418	0.3985584	-1.2435863	0.2136519
provinceNova Scotia	0.8892880	0.1798826	4.9437145	0.0000008
provinceNunavut	1.2232349	1.0880187	1.1242774	0.2608953
provinceOntario	0.6584358	0.1175356	5.6020115	0.0000000
provincePrince Edward Island	1.9463518	0.2962134	6.5707753	0.0000000
provinceQuebec	0.4149517	0.1260342	3.2923740	0.0009935
provinceSaskatchewan	-0.0906261	0.2249622	-0.4028506	0.6870581
educationCollege, CEGEP or other non-university certificate or trade	-0.0133496	0.0772276	-0.1728608	0.8627609
educationHigh school diploma or a high school equivalency certificate	-0.0501428	0.0905801	-0.5535745	0.5798701
educationLess than high school diploma or its equivalent	0.0506284	0.1354018	0.3739121	0.7084697
educationUniversity certificate or diploma below the bachelor's level	0.0344553	0.1012549	0.3402823	0.7336439
educationUniversity certificate, diploma or degree	0.0921349	0.0935233	0.9851542	0.3245483

Figure A5: Descriptive Statistics of the Model of People's Party Regression Estimates

term	estimate	std.error	statistic	p.value
(Intercept)	-4.5709221	0.2432890	-18.7880355	0.0000000
age	-0.6822572	0.1149573	-5.9348766	0.0000000
sexMale	0.6281604	0.1169087	5.3730833	0.0000001
provinceBritish Columbia	0.1956984	0.2407093	0.8130074	0.4162138
provinceManitoba	-0.2225379	0.3685486	-0.6038224	0.5459617
provinceNew Brunswick	0.7529794	0.3590039	2.0974131	0.0359570
provinceNewfoundland and Labrador	-1.2775429	1.0195522	-1.2530431	0.2101900
provinceNova Scotia	0.4592929	0.3464954	1.3255383	0.1849927
provinceNunavut	-12.2904161	900.8230646	-0.0136435	0.9891144
provinceOntario	0.3298977	0.2043245	1.6145770	0.1064023
provincePrince Edward Island	-12.4227714	284.4658695	-0.0436705	0.9651670
provinceQuebec	0.3439537	0.2141060	1.6064643	0.1081719
provinceSaskatchewan	0.3684546	0.3281201	1.1229262	0.2614688
educationCollege, CEGEP or other non-university certificate or trade	0.4458160	0.1652331	2.6981028	0.0069736

term	estimate	std.error	statistic	p.value
educationHigh school diploma or a high school equivalency certificate	0.5746799	0.1815813	3.1648619	0.0015516
educationLess than high school diploma or its equivalent	0.5760795	0.2604136	2.2121712	0.0269548
educationUniversity certificate or diploma below the bachelor's level	0.3235488	0.2154620	1.5016513	0.1331872
educationUniversity certificate, diploma or degree	0.1296214	0.2101867	0.6166967	0.5374348

Figure A6: Descriptive Statistics of the Model of Another Party Regression Estimates

term	estimate	std.error	statistic	p.value
(Intercept)	-5.2584478	0.3631485	-14.4801575	0.0000000
age	0.0422841	0.1669301	0.2533045	0.8000329
sexMale	0.2609277	0.1714604	1.5217960	0.1280602
provinceBritish Columbia	0.4693096	0.3579311	1.3111730	0.1897993
provinceManitoba	0.0274861	0.5343157	0.0514417	0.9589735
provinceNew Brunswick	-0.7736992	1.0427567	-0.7419748	0.4581026
provinceNewfoundland and Labrador	-0.3935838	1.0433762	-0.3772214	0.7060091
provinceNova Scotia	0.6527846	0.5030696	1.2976031	0.1944237
provinceNunavut	-9.4450174	333.3923473	-0.0283300	0.9773989
provinceOntario	0.1373081	0.3257238	0.4215477	0.6733552
provincePrince Edward Island	0.8544655	1.0493388	0.8142894	0.4154792
provinceQuebec	0.4591997	0.3322067	1.3822713	0.1668884
provinceSaskatchewan	1.0561976	0.4203910	2.5124170	0.0119907
educationCollege, CEGEP or other non-university certificate or trade	0.1297466	0.2327176	0.5575283	0.5771665
educationHigh school diploma or a high school equivalency certificate	0.0170259	0.2734136	0.0622716	0.9503466
educationLess than high school diploma or its equivalent	-0.3169694	0.4851169	-0.6533877	0.5135064
educationUniversity certificate or diploma below the bachelor's level	-0.0364350	0.3182725	-0.1144775	0.9088593
educationUniversity certificate, diploma or degree	0.0385057	0.2907081	0.1324547	0.8946247