An Analysis on the Effects of Voter Turnout in the Canadian 2019 Election

Pablo Mercado

12/22/2020

Abstract:

The purpose of this report is to use a multilevel regression model with poststratification (MRP) to showcase the importance and effects of voter turnout, namely in the 2019 Canadian election. An MRP is a statistical technique that takes a multilevel regression model created from survey data and poststratifies the model on census data. Specifically, the MRP was used to make the hypothetical scenario where voter turnout was max capacity during the 2019 Canadian election. The model results are then compared to the real-life results of the election to show the differences and impacts of voter turnout.

Keywords:

Multilevel Regression, Poststratification, Canadian Election, Voter Turnout, Minority Government

Introduction:

One of the most powerful tools every Canadian citizen is given by the Canadian Charter of Rights and Freedoms is the right to vote in an election. The right to vote empowers citizens to have a representative in the state of affairs in the government, effectively allowing citizens to express their ideas and opinions. Despite this, only about two-thirds of eligible voters cast their ballot every election, according to Elections Canada. The number of voters who did not vote is very significant and can have a real impact on the results of every election in Canada's history.

One method to predict how every citizen would vote during an election is by using a multilevel regression model with poststratification (MRP). An MRP is a technique that can be used to weigh out non-representative samples properly and then extrapolate how they would behave. MRP's multilevel regression model uses survey data to analyze the relationship between a response variable and many predictor variables. poststratification is the technique used to predict the model on the target population using census data. In this paper, I will use an MRP to find the 2019 Canadian Federal Election results in the scenario where everyone voted and then see if there will be a significant change in the outcome than what happened in real life.

There are many reasons why eligible Canadians are not voting, but their vote still matters and can heavily influence the formation of the government. This is because Canada's electoral system works where the candidate with the most votes in a riding wins a seat in the House of Commons and represents that riding as a Parliament Member. The Members of Parliament then come together to form the government, and the Prime Minister is the leader of the party with most candidates. If a party does not win the majority of the

seats, the government is known as a minority government. This means that the leading party does not have the singular authority for legislation and must work together with other parties.

For the rest of the report, two data sets will be used for the MRP to showcase what would happen if voter turnout was at max capacity. In the Methodology section, the report describes how a survey data set will be used to create a multilevel regression model and then using that model to poststratify with census data. Results of the model and the poststratification will be in the Results section. And finally, the analysis of the results and conclusions will be presented in the Conclusion section.

Methodology:

Data:

The Canadian Election Study (CES) team created the survey data used for the multilevel regression model; specifically, the model will use the 2019 Online survey. The census data used to poststratify the model was the 2017 General Social Survey (GSS) created by Statistics Canada. The CES made the data set through an online survey during the 2019 federal election campaign, and their goal was to create a record of Canadian society and political life. The GSS data was made with a questionnaire given through the telephone, and its purpose was to monitor changes in Canadian families. The population of interest was all eligible Canadians that can vote, the sampling frame is participants who were asked to take part in either the CES or GSS survey, and the sample are those who filled out either survey.

The Strengths of both of these data sets is that they both managed to collect a large amount of data for each participant. In the CES data set, over 600 variables were created for each participant. In the GSS data set, 81 variables were made for each participant. A weakness for both data sets is that they both have small sample sizes. The CES data set has 30,000 participants, and the GSS data set has around 20,000 participants. This is a weakness because the quantities are considerably smaller than the entire voting population of Canada. This is especially the case with the census data, where it is critical to have many observations.

Both data sets were cleaned in order to conduct the MRP. All missing values are dropped, and the variables of the survey and census data were mutated to match each other. The variables used in both data sets are age, sex, education level, and province, and they are being used as predictor variables for the model. In the survey data, another variable was used: voter choice. It was mutated to create new variables for each party running for the 2019 election to act as the outcome variable. The census data were also aggregated for each variable so that it can be used for poststratification. The descriptive summaries of both data sets can be seen in tables 1 and 2.

Table 1: Desciptive Summary Survey Data

	Overall
n	26344
age (mean (SD))	50.20 (16.68)
sex = Female (%)	14948 (56.7)
educ (%)	
No Highschool Diploma	1272 (6.1)
Highschool Diploma	9695 (46.4)
Undergraduate Degree	6598 (31.6)
Graduate Degree	3345 (16.0)
province (%)	
Alberta	3184 (12.1)
British Columbia	3002 (11.4)
Manitoba	1168 (4.4)
New Brunswick	581 (2.2)
Newfoundland and Labrador	430 (1.6)

	Overall
Nova Scotia	705 (2.7)
Ontario	10490 (39.8)
Prince Edward Island	102 (0.4)
Quebec	5718 (21.7)
Saskatchewan	964 (3.7)

Table 2: Desciptive Summary Census Data

	Overall
n	3945
age (mean (SD))	50.53 (17.90)
sex = Male (%)	1982 (50.2)
educ (%)	
Graduate Degree	720 (18.3)
Highschool Diploma	1195 (30.3)
No Highschool Diploma	861 (21.8)
Undergraduate Degree	1169 (29.6)
province (%)	
Alberta	401 (10.2)
British Columbia	425 (10.8)
Manitoba	374 (9.5)
New Brunswick	369 (9.4)
Newfoundland and Labrador	365 (9.3)
Nova Scotia	396 (10.0)
Ontario	473 (12.0)
Prince Edward Island	306 (7.8)
Quebec	480 (12.2)
Saskatchewan	356 (9.0)

Model:

To determine the results of the 2019 election where all eligible Canadians voted, CES survey data was used to generate seven multilevel logistic regression models, one for each party. Namely, each model's outcome variables are the voter choice for the Liberal Party, the Conservative Party, the NDP, the Bloc Québécois, the Green Party, the People's Party, and Other Parties. Next, individual responses were stratified by province, and then three results: age, sex, and education, were chosen for the independent variables for the models. The reason these responses were chosen was because age and education were shown to be correlated to voter behavior by the American Journal of Political Science, and sex was correlated to voter behavior by Social Science Research.The first variable, age, is in the range of 18 and up. The second variable was sex and split into either Male or Female. The final variable is education, and it was separated into four categories: No Highschool Diploma, Highschool Diploma, Undergraduate Degree, and Graduate Degree. All of these variables are then used for each multilevel logistic regression model. The following is the general equation for each model:

$$\widehat{y} = \log(\frac{p_{vote}}{1-p_{vote}}) = \beta_0 + \beta_1 x_{age} + \beta_2 x_{sex} + \beta_3 x_{education} + \epsilon$$
$$\beta_0 = r_0 + r_1 w_{province} + u$$

This is the general model that will be applied to all parties. the $\widehat{y}$ represents the proportion of Canadian voters who will vote for any given party. Similarly, β_0 represents the intercept of the model and is the

probability of voting for the given party with the categorical variables, sex and education, set to Male and No Highschool Diploma. Furthermore, β_1 represents the slope of the model for each age, β_2 represents the slope of the model for either Male or Female, and β_3 represents the slope of the model for each education level. β_0 is dependent on each province since this is a multilevel regression. Where r_0 is the intercept and r_1 is the slope of the model depending on the province. ϵ and u are the errors.

In order to apply the model to extrapolate the 2019 Canadian election with 100% voter turnout, a post-stratification was performed. This was done by using the results of the model on the GSS census data to find the voting outcome for each group cell, aggregated by province, age, sex, and education. Next, each cell's predicted estimate is weighted by its respective population from the census data. So the results from this poststratification can be used to predict how people in each cell would vote based on their age, sex, and amount of education. The equation for this step is:

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

Where $\widehat{y}^{PS}$ is the proportion of the Canadian population who voted for the given party, N_j represented the number of participants in the cell, and $\widehat{y}_j$ is the cell's predicted vote.

The reason why an MRP was used is because it was necessary to poststratify a model to determine the 2019 Canadian election with 100% voter turnout. There are no other statistical techniques that can accomplish this without going out and surveying all eligible voters, forcing them to pick a party. All the models were made with the frequentist approach, and the warnings of large eigenvalues were shown for the unpopular parties like the Bloc Québécois and the Green Party.

Results:

Table 3: Summary of Model for the Liberal Party

Variable	Coefficients	P.values
Intercept	-1.3864689	6.38e-15
Age	0.0051179	9.48e-09
Sex: Female	0.0989987	0.00118
Education: Highschool Diploma	0.1847655	0.00711
Education: Undergraduate Degree	0.5741996	$<$ 2e-16
Education: Graduate Degree	0.6375582	$<$ 2e-16

Table 4: Summary of Model for the Conservative Party

Variable	Coefficients	P.values
Intercept	-1.2196286	1.89e-08
Age	0.0141887	$<$ 2e-16
Sex: Female	-0.3794575	$<$ 2e-16
Education: Highschool Diploma	0.0438191	0.5103
Education: Undergraduate Degree	-0.1529645	0.0259
Education: Graduate Degree	-0.3574165	1.83e-06

Tables 3 and 4 are a summary of the model for the Liberal and Conservative Parties. The tables for the other parties are in the appendix. The Variable column describes the variable's name and the factors for each categorical variable. The coefficients describe the β values for the model described above. And finally,

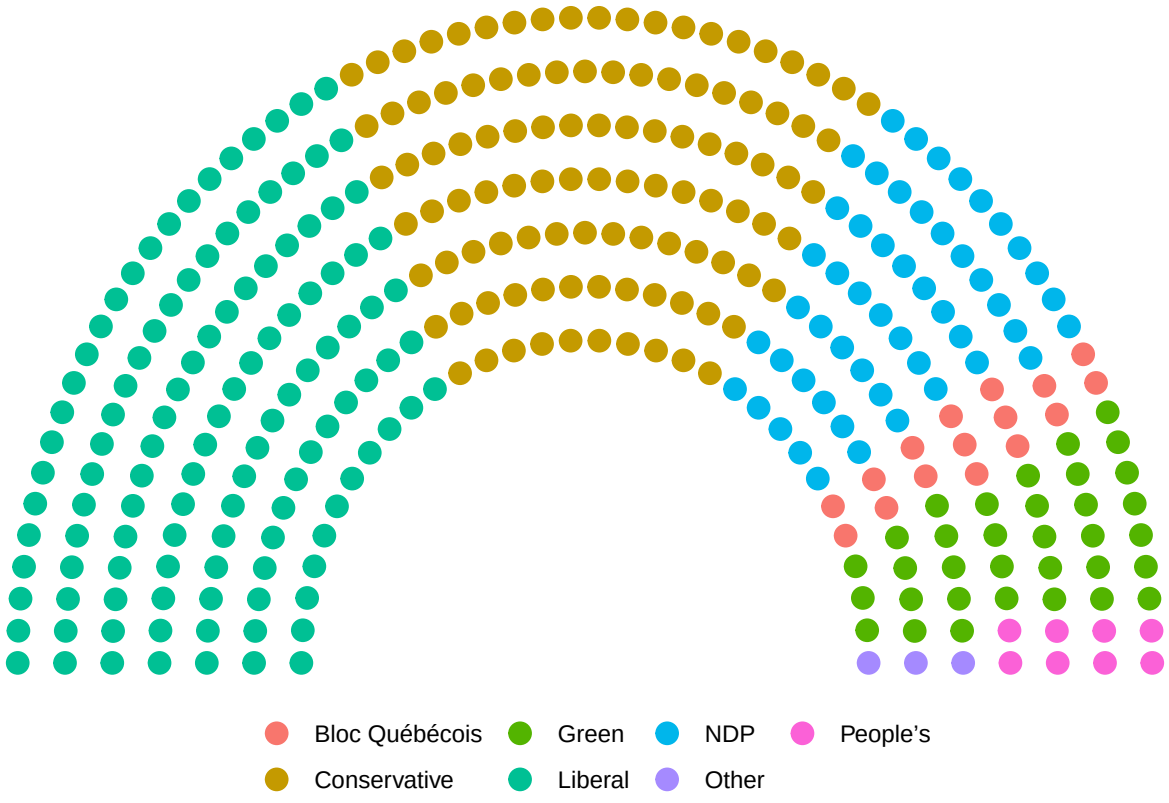
the p-values describe the statistical significance of each variable for the model. In each model, most variables were extremely significant with a $P < 0.01$. All of these values were created using the frequentist multilevel regression with the binomial family for a logistic equation.

Table 5: Proportion of Party Votes at 100% Voter Turnout

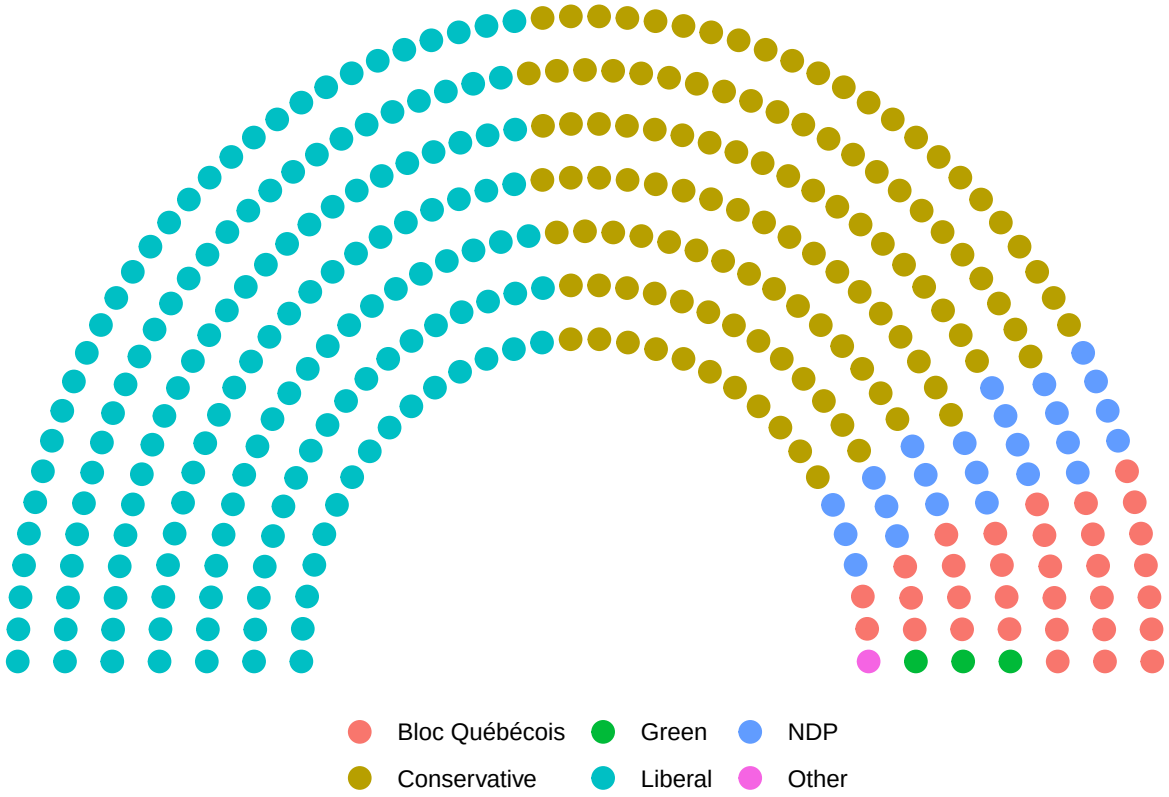
Party	Population.Proportion.Votes
Liberal Party	0.351
Conservative Party	0.314
NDP	0.153
Bloc Québécois	0.048
Green Party	0.100
People's Party	0.023
Other	0.011

Table 5 shows all the proportion of votes that each party would get at max capacity voter turnout. This was done by poststratifying the models for each party, which accounted for the variables province, age, sex, and education.

Graph 1: Predicted Seats of Canadian Parliament at 100% Voter Turnout



Graph 2: Real Seats of Canadian Parliament from the 2019 Canadian Election



Graph 1 is a graphical representation of the seats won by each party if voter turnout was at 100%. This was created by taking the population proportion of votes and converting them into the proportion of 338 seats. Graph 2 is real seating results which the data was taken from Elections Canada.

Discussion:

Summary:

As a quick synopsis, this paper tried to find the importance and effects of full voter turnout in the 2019 Canadian Federal election. This was done by performing an MRP to create a prediction on the proportions of the population that would vote for each party. A multilevel logistic regression model was created for each party running for parliament using the CES data set to act as the survey data for the model. The Model is using age, sex, and education as level 1 predictor variables and the province as a level 2 predictor variable. The independent variable for the model is a logistic choice for voting for a given party. The models are then poststratified with the GSS data set, which acted as the census data. After the Calculations, tables were created to describe the models, and a graph was created to showcase what the House of Commons' seats would look like.

Conclusion:

Firstly, the models mostly showed that the variables chosen were statistically significant. This is because the p-values of the variables are below 0.05. The variable that was most significant across all models was age. This means that age is a good indicator of how an Eligible Canadian would vote. The variable that would sometimes not be significant was education. This means that education is not a reliable measure of

voter behavior. This was especially seen in the models for the Green Party and the Conservative Party. For the poststratification step, the results of applying the model to the GSS data can be seen in Table 5.

Graph 1, which was created by the predicted results of the MRP, and graph 2, which was created using the real results of the 2019 election, can be compared to find the effects of voter turnout. Visually, the 2 bigger parties, liberal and conservative, have less representation, and the seats are more spread out to other parties. Table 6 shows the difference in seats from predicted to real results.This means that voter participation clearly affects the results of the 2019 Canadian election. If the predicted House of Commons were real, this would mean the liberal and conservative parties would have less power and that they would need to work together with other parties to pass legislation. In conclusion, voter turnout is one of the most important factors for any election, and it is important for voters to vote so that they can be properly represented in Government.

Table 6: Party Seating in the House of Commons

Party	Predicted.Seats	Real.Seats
Liberal Party	119	157
Conservative Party	106	121
NDP	52	24
Bloc Québécois	16	32
Green Party	34	3
People's Party	8	0
Other	3	1

Weaknesses:

One weakness that this analysis had was in the model. As it was found, the p-value results for education and sex were not always statistically significant and had p-values greater than 0.05. This means that these variables were not reliable, especially compared to age which was significant across all models. More weaknesses are in the data sets used. First, the GSS size is way too small, containing only around 20,000 observations, and is not a good census data that should be representative of the eligible voting population of Canada. One final weakness from the is the fact that the GSS did not include the territories of Canada, so it was impossible to have their representation in the model. This would greatly affect the study since the topic of interest was all eligible voters in Canada, not just in the provinces.

Next Steps:

The next steps of this paper are to improve the model and data sets used to fix the difficulties described in the Weaknesses section. The first step is to change the variables of sex and education in the model. According to Social Science Research, two good indicators for voter behavior is religious affiliation and wage. The next steps for the data set would be to use proper census data, such as the census created by Statistics Canada. This will likely fix the sample size issue and include those from provinces.

References:

Elections Canada. (2013, January 14). Retrieved December 10, 2020, from https://www.elections.ca/

Azzi, S. (2012, January 17). Minority Governments in Canada. Retrieved December 22, 2020, from https://www.thecanadianencyclopedia.ca/en/article/minority-government

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

Government of Canada, Statistics Canada. Population Estimates on July 1, by Age and Sex, Government of Canada, Statistics Canada, September 29, 2020, www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1710000501.

Brooks, C., Nieuwbeerta, P., & Manza, J. (2006). Cleavage-based voting behavior in cross-national perspective: Evidence from six postwar democracies. Social Science Research, 35(1), 88-128. doi:10.1016/j.ssresearch.2004.06.005

Bartels, L. M. (2000). Partisanship and Voting Behavior, 1952-1996. American Journal of Political Science, 44(1), 35. doi:10.2307/2669291

Leeper, T. J. (n.d.). Ggparliament. Retrieved December 22, 2020, from https://www.rdocumentation.org/packages/ggparliament/versions/0.1.2/topics/ggparliament

2019 Canadian federal election. (2020, December 17). Retrieved December 22, 2020, from https://en.wikipedia.org/wiki/2019_Canadian_federal_election

Alexander, R; Caetano, S. (2020, October 7). gss_cleaning.R (Version 1) [Source code]. University of Toronto.

Appendix

Table 7: Summary of Model for the Green Party

Variable	Coefficients	P.values
Intercept	-1.6518959	5.62e-11
Age	-0.0132158	< 2e-16
Sex: Female	0.1787582	0.000361
Education: Highschool Diploma	-0.0001664	0.998722
Education: Undergraduate Degree	-0.0865395	0.419506
Education: Graduate Degree	0.1149236	0.312166

Table 8: Summary of Model for the People's Party

Variable	Coefficients	P.values
Intercept	1.873743	< 2e-16
Age	-0.020562	2.36e-13
Sex: Female	-0.558684	4.96e-09
Education: Highschool Diploma	-0.501494	0.000972
Education: Undergraduate Degree	-1.014754	2.26e-09
Education: Graduate Degree	-0.810526	1.48e-05

Table 9: Summary of Model for the NDP

Variable	Coefficients	P.values
Intercept	-0.532480	0.000579
Age	-0.026640	< 2e-16
Sex: Female	0.458265	< 2e-16
Education: Highschool Diploma	-0.120774	0.129614
Education: Undergraduate Degree	-0.241970	0.003292
Education: Graduate Degree	-0.190717	0.034056

Table 10: Summary of Model for the Bloc Québécois

Variable	Coefficients	P.values
Intercept	-17.96000	<2e-16
Age	0.02569	< 2e-16
Sex: Female	-0.04817	< 2e-16
Education: Highschool Diploma	-0.31010	<2e-16
Education: Undergraduate Degree	-0.55560	<2e-16
Education: Graduate Degree	-0.57870	<2e-16

Table 11: Summary of Model for Other Parties

Variable	Coefficients	P.values
Intercept	-5.752630	< 2e-16
Age	0.013137	0.00609
Sex: Female	-0.228096	0.15948
Education: Highschool Diploma	0.458005	0.24549
Education: Undergraduate Degree	0.160085	0.69577
Education: Graduate Degree	0.259967	0.54458