Mandatory Voting Laws May Change Outcomes of Elections, Examining Voter Turnout in the 2019 Canadian Election

Jessica Glustien

December 20, 2020

Abstract

The importance of voter turnout is a topic that has long been debated. The 2019 Canadian federal election was a tight race with only 77% voter turnout (Elections Canada, 2019). Through multi level regression models and the use of post-stratification, this paper examines the demographic trends of voting and generates a prediction of the 2019 election results if all eligible voters had participated. Ultimately it is not clear whether the winner would be altered by this change, but a significant increase in popular vote margin between the Conservative party and the Liberal party is observed. Code and data supporting this analysis are available at https://github.com/jessglustien/sta304-final (https://github.com/jessglustien/sta304-final)

Keywords

Post Stratification, Multiple Regression Model, Survey, Elections, Voter Turnout

Introduction

Federal elections are the process through which Canadians decide the future of our country. Despite such an impactful result, many Canadians have a lack of interest in the electoral process and are unmotivated to go out and vote. In the federal election of 2000, only 61 percent of eligible voters cast a ballot, the lowest voter turnout rate since Confederation. (Blais et al., 2002) Assuming that many of that missing 39 percent lean in some political direction, this allows for a large margin of error between the final election result and the true opinion of all citizens.

There has been a large amount of research done into the effectiveness and feasibility of mandatory voting laws, and it has been found that "if the electorate is sufficiently large, then increasing voter turnout is generically efficient." (Krasa & Polborn, 2009). This leads to the question of whether a tangible benefit would be found by implementing such a law. While we know that the results would take into consideration a greater number of Canadian's opinions, this paper will examine the probability that having all eligible voters cast a ballot will change the result of elections. If it is unlikely that election results will change, then it is harder to justify the cost of enforcing mandatory voting. If there is a significant likelihood of a different result, this could help support an argument for mandatory voting laws.

This examination will be done by completing a multi-level regression on the Canadian Election Study data collected during the 2019 federal election, and then using post stratification with the General Social Survey data collected by Statistics Canada in 2017 to predict the outcome of the 2019 election if all Canadians had voted.

Methodology

Data

Two different datasets will be used in this paper. The first dataset is the Canadian Election Study data that contains information about voting intention for the 2019 federal election. This data will be used to construct a model that can predict the likelihood of voting for a certain party, based on a person's age, gender, education level, and the province that they live in. This data was collected through an online survey conducted through Qualtrics, and targets where stratified by region and balanced on gender and age. The target population was all eligible voters in the 2019 election, so Canadian citizens and permanent residents over the age of 18. Over 30,000 responses were received over the course the survey. (Stephenson et al., 2020)

The second dataset is the General Social Survey data collected by Statistics Canada. This is the data that will be used for the post stratification process. It was chosen to be a more accurate representation of the Canadian population in comporasion to the CES data. This data was collected by Random Digit Dialing and the digital survey is completed by respondents using Computer Assisted Telephone Interviewing. The target population was all Canadians over the age of 15, and over 25,000 responses were collected. (Statistics Canada, 2017) In order for this data to be used to predict voting outcomes, it was necessary to exclude responses from people under the age of 18, as they are not considered eligible voters. The strength of this dataset is that the randomized approach to collect responses allows for it to be representative of the country's demographics, and so is a good candidate for use in post-stratification. The weak point of this dataset is that only ~25,000 responses were collected, so while those responses are spread proportionally over demographics, a larger number of responses such as a census would be better for post-stratification.

Both datasets contain information about the age, gender, education level, and home province of the respondents. This will allow for the model created with the CES data to be applied to the GSS data. One of the decisions made when cleaning the data for this project was the create a new variable age_as_factor, that places each respondent into a cell depending on what range their age falls in. This choice was made because in the post stratification step, we will be sorting respondents into boxes depending on each variable value, and converting age from a continuous variable into a categorical one will allow for more significant boxes. (As otherwise, many boxes would only contain one person since age is measured to such specificity). Age, gender, and education were chosen for the model as these demographics have been shown to be tied to voting habits. For example, it has been observed that "well-educated citizens vote more frequently than the poorly educated" (Gallego, 2010) and that women are "leaning more to the left in their voting decisions" (Abendschön & Steinmetz, 2014) when comparted to their male counterparts.

When creating the models the CES data has the following important attributes that will be used as the binary outcome variables.

Variable	**= 1**	**= 0**
vote_liberal	Intends to vote for Liberal Party	Intends not to vote for Liberal Party
vote_conservative	Intends to vote for Conservative Party	Intends not to vote for Conservative Party
vote_ndp	Intends to vote for NDP	Intends not to vote for NDP

This will allow us to predict the popular vote proportions that will be received by each party. We can also view the count of intended votes of respondents in the CES data.

This shows the support between the Liberal party and the Conservative party was very close going into the election. When viewing the popular vote of the completed election, the Conservatives actually received slightly more votes than the Liberal party, despite this graph showing the opposite relationship. (Elections Canada, 2020) This could imply that there was a selection bias in this survey that caused Liberal voters to be more likely to respond. This weakness would be compensated for by the post-stratification. Another interesting observation is that despite the NDP being the third most popular party in the country, there is actually a larger proportion of support for a smaller “Other” party than for the NDP.

Multi-Level Regression

A multi-level regression model will be used to predict the voting habits of a person as determined by their age, gender, education, and home province. A multi-level model was chosen over a traditional regression model to help increase the validity of the inference and account for group trends. In typical regression, all observations are treated as independent, but when categories that apply on the group level, such as home province, this assumption is no longer correct. There may be correlation between voting habits of people living in a certain province. Multi-level regression accounts for this distinction by first creating a model at the individual level (accounting only for age, gender, and education), and then creating a second model at the group level that will be used to determine the intercept of the first model.

In this paper we will be creating three logistic multi-level regression models. Each will be used to predict the proportion of votes a particular party would receive if all eligible voters had voted. Each model will represent one party; liberal, convervative, and NDP. As these three parties have historically received a large majority of the population’s vote (Campbell & Christian, 1989), it is these proportions that will decide the outcome of the election. Our models will take the following form:

Table 2: Model Breakdown

Model	Outcome Variable	Description
Model I	vote_liberal	The predicted proportion of votes the liberal party would receive if all eligible voters participated.
Model II	vote_conservative	The predicted proportion of votes the convervative party would receive if all eligible voters participated.
Model III	vote_ndp	The predicted proportion of votes the NDP would receive if all eligible voters participated.

All three models are created from the CES data, and will take on the following mathematical form:

Level 1:

$$\log\left(\frac{\hat{Y}}{1-\hat{Y}}\right) = \hat{\beta}_{0j} + \hat{\beta}_1 x_{1j} + \hat{\beta}_2 x_{2j} + \hat{\beta}_3 x_{3j} + \hat{\beta}_4 x_{4j} + \hat{\beta}_5 x_{5j} + \hat{\beta}_6 x_{6j} + \hat{\beta}_7 x_{7j} + \hat{\beta}_8 x_{8j} + \hat{\beta}_9 x_{9j} + \hat{\beta}_{10} x_{10j}$$

Where $\hat{Y}$ is the predicted probability of the person voting for the respective party. In this equation the individual variables are multiplied by the same constants (β_1-β_{10}) regardless of group level (province), but the intercept value $beta_0$ changes depending on province. This is represented through the index j, which indicates the province the data is from. The mathematical model for β_0 is:

Level 2:

$$\hat{\beta}_{0j} = r_{00} + r_{0j} W_j$$

Where r_{0j} is the slope dependant on the province.

Table 3: Breakdown of Variable Categories

Notation	Variable	Category	Reference Level
x_1	Age	21-40	0-20
x_2	Age	41-60	0-20
x_3	Age	61-80	0-20
x_4	Age	81+	0-20
x_5	Gender	Male	Female
x_6	Education	Bechelor's Degree	Masters or higher
x_7	Education	College	Masters or higher
x_8	Education	Highschool	Masters or higher
x_9	Education	Less than Highschool	Masters or higher
$x_1 0$	Education	Other	Masters or higher
W_j	Province	–	–

Post-Stratification

Once the models have been created, we can predict the proportion the respective party will receive using post stratification. This process involves using demographic information found in the CES dataset to extrapolate how the entire population vote by determining the proportion of each demographic and giving the appropriate weight to their predicted voting trends. This technique uses the following model:

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

What this equation is doing is breaking our data into cells based off of the demographic information our model uses. So each cell is split up by age, gender, education, and province, and with each response falling into exactly one cell. We're then summing up our predicted voting proportion for each cell, and finding a weighted average. Each cell is weighted according to how the cell's demographic is represented overall in Canada. This is why it is so important that the GSS data we are using for post stratification is a representative sample of the population.

Results

Table 4: Model 1 - Vote Liberal

	as.factor(vote_liberal)		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.61	0.45 – 0.83	**0.002**
age_factor21-40	0.80	0.69 – 0.94	**0.005**
age_factor41-60	0.80	0.68 – 0.93	**0.003**

age_factor61-80	0.95	0.82 – 1.11	0.532
age_factor [81+]	0.92	0.71 – 1.19	0.513
gender [Male]	1.01	0.96 – 1.06	0.774
education [Bachelor's Degree]	0.89	0.82 – 0.96	**0.003**
education [College]	0.60	0.55 – 0.66	**<0.001**
education [Highschool]	0.57	0.53 – 0.62	**<0.001**
education [Less than highschool]	0.53	0.46 – 0.61	**<0.001**
education [Other]	0.45	0.24 – 0.85	**0.014**
Random Effects			
σ^2	3.29		
$\tau_{00\ province}$	0.20		
ICC	0.06		
$N_{province}$	13		
Observations	31564		
Marginal R^2 / Conditional R^2	0.017 / 0.074		

Table 5: Model 2 - Vote Conservative

	as.factor(vote_conservative)		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.10	0.07 – 0.15	**<0.001**
age_factor21-40	1.79	1.48 – 2.16	**<0.001**
age_factor41-60	2.38	1.98 – 2.87	**<0.001**
age_factor61-80	2.50	2.08 – 3.02	**<0.001**
age_factor [81+]	3.01	2.28 – 3.98	**<0.001**
gender [Male]	1.64	1.56 – 1.73	**<0.001**
education [Bachelor's Degree]	1.16	1.06 – 1.27	**0.001**
education [College]	1.47	1.34 – 1.62	**<0.001**
education [Highschool]	1.56	1.43 – 1.71	**<0.001**
education [Less than highschool]	1.38	1.20 – 1.59	**<0.001**

education [Other]	1.06	0.59 – 1.92	0.840
Random Effects			
σ^2	3.29		
$\tau_{00\ province}$	0.36		
ICC	0.10		
$N_{province}$	13		
Observations	31564		
Marginal R^2 / Conditional R^2	0.035 / 0.130		

Table 6: Model 3 - Vote NDP

	as.factor(vote_ndp)		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.33	0.25 – 0.45	**<0.001**
age_factor21-40	0.81	0.69 – 0.95	**0.009**
age_factor41-60	0.43	0.36 – 0.51	**<0.001**
age_factor61-80	0.28	0.23 – 0.33	**<0.001**
age_factor [81+]	0.24	0.16 – 0.36	**<0.001**
gender [Male]	0.69	0.64 – 0.74	**<0.001**
education [Bachelor's Degree]	1.02	0.91 – 1.14	0.717
education [College]	1.06	0.94 – 1.20	0.323
education [Highschool]	1.05	0.93 – 1.18	0.436
education [Less than highschool]	1.22	1.03 – 1.45	**0.021**
education [Other]	0.88	0.42 – 1.80	0.717
Random Effects			
σ^2	3.29		
$\tau_{00\ province}$	0.16		
ICC	0.05		
$N_{province}$	13		
Observations	31564		
Marginal R^2 / Conditional R^2	0.069 / 0.111		

Table 7: Summary of Post-Stratified Model Results

Model	Outcome Variable	Predicted Outcome (Popular Vote)
Model I	Vote Liberal	21.9%
Model II	Vote Conservative	25.8%
Model III	Vote NDP	11.8%

After completing our post-stratification analysis of predicted vote proportion for the liberal party, conservative party and NDP, using a multi-level regression model, which accounted for age, gender, education level, and home province, we have determined that if voter turnout was 100% in the 2019 Canadian federal election, the liberal party would receive 21.9% of the popular vote, the conservative party would receive 25.8% of the popular vote, and the NDP would receive 11.8% of the popular vote.

Discussion

Summary

In order to complete any meaningful analysis of these values, we must be able to compare them to the actual popular vote result of the 2019 election. Using the official voting results from Elections Canada, we can create the following comparison:

Table 8: 2019 Federal Election Popular Vote

Party	Popular Vote	Predicted Popular Vote
Liberal	33.1%	21.9%
Conservative	34.3%	25.8%
NDP	19.7%	11.8%

(Elections Canada, 2019)

The most apparent takeaway is that all of the predicted proportions are significantly lower than the results of the election. The overall order between the parties remains the same, with the Conservative party gaining a larger lead over the Liberals in the projected results. When interpreting the created models it is also important to note that the tables display the odds ratios of the different variables, instead of the coefficients shown in the mathematical model. The odds ratios are created by taking the probability of success (voting for the party) over the probability of failure. This means an odds ratio of greater than one implies a greater change of voting for the party, while an odds ratio below one implies the opposite. An odds rario of close to 1 implies the variable has little affect on the voting choice.

When examining the created models it is interesting to note that age and gender have the strongest correlation for whether an individual will vote for the NDP. Age appears to be negatively correlated with voting NDP, and women are more likely than men to vote NDP. Liberal voters do not seem as strongly split along these categories. There seems to be almost no correlation between gender and voting Liberal. The Conservative model has low p-values for all variables except the Other category of education, showing that all the investigated demographics have a correlation with whether an individual will vote Conservative. Older voters have very large odds ratio showing that they are more likely to vote Conservative than younger voters.

Conclusions

To return to the motivating question of this paper, do these results imply that full voter turnout would affect the results of the election? There is not a definitive answer to that question, as the weakness of this model is that it predicts popular vote, but the election is decided by ridings. The predicted results do show a much larger margin between the Conservative party and the Liberal party, this increase in margin could very well be enough to change the winner from the Liberals to the Conservatives, depending on the geographic spread of the additional Conservative votes.

The other important result that is shown by this prediction is the decrease in popular vote across all three of Canada's largest political parties. This implies that a large percent of people who didn't vote in 2019, would vote for one of the less popular parties. A possible motivating factor of this could be that feeling your supported party has no chance of winning, you are less likely to be motivated to vote.

Both of these conclusions offer support to enforcing mandatory voting. The predicted results are significantly different from the actual results, that there is a possibility that mandatory voting would results in a different election outcome. It can also be said that if supporters of smaller parties are unlikely to show their support through voting, these smaller parties will never have the opportunity to grow. By enforcing mandatory voting, these parties will be given a chance to have their voices heard.

Weaknesses and Next Steps

The largest weakness of this study was the fact that it only examined trends within the popular vote, and not predicting the outcomes for each riding. This makes it impossible to determine which party would definitively win the election under our model. While many interesting and valuable can still be made from the popular vote predictions, going forward it would be a significant extension to this study to use detailed riding and geographic information to complete the same prediction but at a riding level.

Another weakness is that while the GSS data is a good representation of the Canadian voting population, a more accurate post-stratification could be completed if actual census data was used. This would ensure that the demographic information and ratios being used in the post-stratification process accurately reflected the country. It is also important to note that the GSS data was collected in 2017, while the CES data is from 2019. While these are close enough together that the demographic changes would not be large, it would still be a stronger prediction if the post-stratification data was also taken in 2019.

Looking forward it would be interesting to delve further into the noticed correlation between an individual not voting and the increased liklihood that said individual supports a smaller political party (ie. not the Liberals, Conservatives, or NDP). This could be done by completing a regression model between whether an individual participates in federal elections and the binary outcome of if they support a smaller political party or not. Even further, it would be interesting to gain insight through a survey as to the reasons why people choose not to vote, to help either support or disprove the hypothesis that many are not voting because they feel their party does not stand a chance of winning.

References

Abendschön, S., & Steinmetz, S. (2014). The gender gap in voting revisited: Women's party preferences in a European context. Social Politics, 21(2), 315-344.

Blais, A., Gidengil, E., Nadeau, R., & Nevitte, N. (2002). Generational Change and the Decline of Political Participation: the case of voter turnout in Canada. Citizenship on Trial: Interdisciplinary Perspectives on the Socialization of Adolescents, Montreal.

Campbell, C., & Christian, W. (1989). Political parties and ideologies in Canada. McGraw-Hill Ryerson.

Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01 (doi:10.18637/jss.v067.i01).

"Forty-Third General Election 2019: Official Voting Results". Elections Canada. October 22, 2019.

Gallego, A. (2010). Understanding unequal turnout: Education and voting in comparative perspective. Electoral Studies, 29(2), 239-248.

Krasa, S., & Polborn, M. K. (2009). Is mandatory voting better than voluntary voting?. Games and Economic Behavior, 66(1), 275-291.

Lüdecke D (2020). *sjPlot: Data Visualization for Statistics in Social Science*. R package version 2.8.6, <URL: https://CRAN.R-project.org/package=sjPlot (https://CRAN.R-project.org/package=sjPlot)>.

Statistics Canada (2017, December 19). General Social Survey - Family (GSS). Government of Canada. https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&Id=335816 (https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&Id=335816) (https://www23.statcan.gc.ca/imdb/p2SV.pl (https://www23.statcan.gc.ca/imdb/p2SV.pl)?Function=getSurvey&Id=335816)

Stephenson, Laura B; Harell, Allison; Rubenson, Daniel; Loewen, Peter John, 2020, "2019 Canadian Election Study - Online Survey", https://doi.org/10.7910/DVN/DUS88V (https://doi.org/10.7910/DVN/DUS88V), Harvard Dataverse, V1

Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686 (https://doi.org/10.21105/joss.01686)