

Analysis on 2019 Canadian Federal Election: What would be different if everyone had voted?

Yongpeng Hua 1003865846

09/12/2020

Data input

Load in both survey data and census data.

Abstract

The following report is about the analysis on the assumption that how the voting result would change if people who didn't vote all vote to the Liberal Party.

Keywords

Vote, First choice, Liberal Party

Introduction

The 2019 Canadian federal election (formally the 43rd Canadian general election) was held on October 21, 2019, to elect members of the House of Commons to the 43rd Canadian Parliament. The Liberal Party, led by incumbent Prime Minister Justin Trudeau, won 157 seats to form a minority government and lost the majority they had won in the 2015 election. The Liberals lost the popular vote to the Conservatives, which marks only the second time in Canadian history that a governing party formed a government while receiving less than 35 per cent of the national popular vote. (Wikipedia, 2019) However, when looking at the dataset of 2019 Canadian Federal Election, it is necessary to mention that there exist a lot of people who didn't vote to any party. If everyone who didn't vote had voted for their first choice, how this would affect the results of the election? Will Liberal party still win the election? In other words, if everyone voted to their first choice, how many of them would vote for Liberal party?

One way to solve these puzzles is to produce an analysis with MRP method, which is a method relied on a statistical method called "post-stratification". Post-stratification is a technique for adjusting a non-representative sample for which there are demographic predictors characterizing the strata. It is carried out after a model is fit to the observed data, hence the name post-stratification (Little, 1993). Multilevel modeling is useful in determining how much partial pooling to apply in the regressions, leading to the popularity of the combination of multilevel regression and post-stratification (MRP) (Park, Gelman, and Bafumi, 2004).

Two datasets will be used in this analysis. One is called "2019 Canadian Election Study - Online Survey", and the other one is called "Canada census data". In the Methodology section, I will describe these data in detail and the model that was used to perform the analysis. Results of the analysis are presented in the Results section, and conclusions & inferences of the data are presented in the Conclusion section.

Methodology

Data

To process the analysis, I cleaned up both survey data and census data first. These data are saved as “survey_data.csv” and “census_data.csv”.

Model

```
attach(survey_data)
# Build the model
logitm <- glm(vote_liberal ~ as.factor(sex) + as.factor(age) + as.factor(education))
detach(survey_data)
```

Post-Stratification

```
census_data$logodds_estimate <-
  logitm %>%
  predict(newdata = census_data)

census_data$estimate <-
  exp(census_data$logodds_estimate)/(1+exp(census_data$logodds_estimate))

yps <- census_data %>%
  mutate(alp_predict_prop = estimate*n) %>%
  summarise(alp_predict = sum(alp_predict_prop)/sum(n))
yps
```

Results

Discussion

Summary

Conclusion

Weakness & Next Steps

References

Please refer to the file “reference.bib”.