Canadian Federal Election analysis focus on the different political spectrum

Author: zhijian zhu(1002943201)
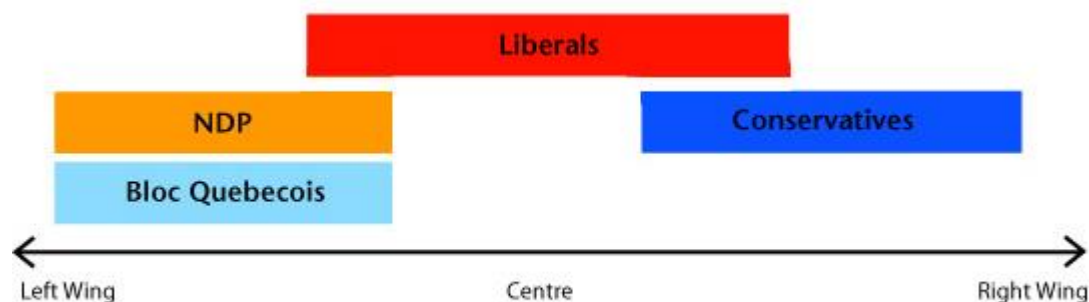
Due Date: Dec. 22, 2020

Abstract

I would like to apply it to the Canadian federal election now that we have predicted the popular vote in the US federal election in 2020. I have built a new model by using the data from the sub dataset of CESR, the decon. Also, since I could not apply the multivariate post-stratification method on the Canadian census dataset from the "Statistics Canada" website. Which I have done the prediction by using the GSS data instead of CES data.

Key word

Canadian Federal Election, turnout rate, political spectrum is look like this:

[1]

---

[1]

http://bclearningnetwork.com/LOR/media/SS11/COURSE_4034644_M/my_files/module1/section1/lesson2/topic2.html, 2006, Open School BC

Introduction:

The federal election has always been a national priority because it will determine the appointment and removal of the prime minister and affect the government's future policies. I plan to construct a multivariate Logistic (Logit) regression model by using a series of individual level variables: age, gender, first language, education level, income, marital status, and a group level variable: province. There is no column indicate the federal party that the candidate voted in the dataset of decon, then I decided to use the political spectrum as the response variable. Then use the political spectrum to find out the most similar party as the election result. After getting the model, I will use the GSS data to predict the political spectrum and find out the most similar federal party corresponding to it, for example the Liberal Party of Canada would in the center and Conservative Party of Canada would in the right. I support the result will not in the center which is different from the result of 2019 Canadian federal election since we assume turnout rate is 100%.

Data:

I clean the data learn from the gss_cleaning.R from Prof. Rohan alexander. In decon data,

I transfer the income to the income level which aim to match the GSS data.

Model:

First model:

$$\log\frac{p}{1-p} = \beta_0 + \beta_1 \text{age} + \beta_2 \text{gender} + \beta_3 \text{education} + \beta_4 \text{income} + \beta_5 \text{province}$$

Second model:

$$\text{political spectrum} = \beta_0 + \beta_1 \text{age} + \beta_2 \text{gender} + \beta_3 \text{education} + \beta_4 \text{income} + \beta_5 \text{province}$$

where,

1. p represents the probability of voters who will vote for right wings

2. age term is the age of indicate people

3. gender term is the gender of indicate people

4. education term is the highest education achievement of indicate people

5. income term is the income level of indicate people

6. province term is the province indicate people live in

First model is using the logistic model and second model is the linear model. Since logistic model may not represent the political spectrum well, then the second linear model may be better in this case.

The province is the group level for the Post-Stratification method and the rest of terms are the individual levels.

Results:

First model:

The summary table for logistic model

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.251	0.082	3.066	0.002
age	0.001	0.001	1.610	0.107
sexMale	0.460	0.030	15.514	0.000
education	-0.111	0.016	-7.101	0.000
provinceBritish Columbia	-0.556	0.060	-9.224	0.000
provinceManitoba	-0.343	0.081	-4.250	0.000
provinceNew Brunswick	-0.436	0.106	-4.119	0.000
provinceNewfoundland and Labrador	-0.345	0.127	-2.720	0.007
provinceNorthwest Territories	-1.842	0.773	-2.382	0.017
provinceNova Scotia	-0.828	0.105	-7.859	0.000
provinceNunavut	-0.297	0.544	-0.546	0.585
provinceOntario	-0.456	0.048	-9.469	0.000
provincePrince Edward Island	-1.062	0.268	-3.960	0.000
provinceQuebec	-0.567	0.052	-10.835	0.000
provinceSaskatchewan	-0.251	0.088	-2.847	0.004
provinceYukon	-0.442	0.419	-1.054	0.292
income_respondent$125,000 and more	0.062	0.054	1.154	0.249
income_respondent$25,000 to $49,999	-0.270	0.056	-4.870	0.000
income_respondent$50,000 to $74,999	-0.118	0.053	-2.239	0.025
income_respondent$75,000 to $99,999	0.016	0.054	0.301	0.763
income_respondentLess than $25,000	-0.268	0.063	-4.234	0.000

We can see the age terms are not significant and most levels in sex, education, province, and income level have different political tendency.

Point estimate and margin of error of probability of vote for right wing

Point.estimate	Margin.of.error
0.4206	0.1741

The estimate result by using GSS data is 42.06% people will vote for **right wing.** That is same as the real result of 2019 Canadian federal election - the Liberal Party, led by incumbent Prime Minister Justin Trudeau, won 157 seats to form a minority government.

Second model:

The summary table for linear model

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	5.682	0.097	58.836	0.000
age	0.008	0.001	7.598	0.000
sexMale	0.559	0.035	16.026	0.000
education	-0.242	0.018	-13.203	0.000
provinceBritish Columbia	-0.720	0.071	-10.156	0.000
provinceManitoba	-0.355	0.096	-3.711	0.000
provinceNew Brunswick	-0.529	0.125	-4.240	0.000
provinceNewfoundland and Labrador	-0.260	0.150	-1.730	0.084
provinceNorthwest Territories	-1.295	0.667	-1.941	0.052
provinceNova Scotia	-0.923	0.120	-7.717	0.000
provinceNunavut	-0.248	0.643	-0.385	0.700
provinceOntario	-0.521	0.057	-9.171	0.000
provincePrince Edward Island	-1.166	0.295	-3.948	0.000
provinceQuebec	-0.656	0.062	-10.658	0.000
provinceSaskatchewan	-0.237	0.105	-2.262	0.024
provinceYukon	-0.512	0.492	-1.040	0.298
income_respondent$125,000 and more	0.059	0.064	0.924	0.356
income_respondent$25,000 to $49,999	-0.321	0.065	-4.931	0.000
income_respondent$50,000 to $74,999	-0.164	0.062	-2.639	0.008
income_respondent$75,000 to $99,999	0.011	0.064	0.171	0.864
income_respondentLess than $25,000	-0.268	0.074	-3.622	0.000

We can see almost all the terms are significant which means most levels in age, sex, education, province, and income level have different political tendency.

Point estimate and margin of error of probability of vote for right wing

Point.estimate	Margin.of.error
5.1544	1.0232

The estimate result of political spectrum would be 5.15, which is quite similar to the situation of 2019 Canadian federal election. The Liberal Party and the Conservative Party have close electoral votes. The Liberal Party of Canada is the center-left wing and Conservative Party of Canada is the right wing. Therefore, the result of political spectrum in real 2019 Canadian federal election should be in the center-right wing which is same as my estimate result.

Summary:

The two model estimate the result from different model, and they show a different result. However, after comparing the result, the result just estimates the 2019 Canadian federal election from the different perspective and they both got the similar result to the real 2019 Canadian federal election

Conclusions:

The result is similar to the real 2019 Canadian federal election, then it is hard to say the result would be different if everyone votes in the election. However, the result of probability of vote for right wing close to 50% and the political spectrum close to 5 which means the election is intense and each vote is important for the win.

Weakness & Next Steps:

The response variable is not directly estimating the probability of winning that is hard to see the result directly and make the estimate unreliable. Next steps will be trying to make the response variable to be the probability of winning.

Because the data portion is coming from the mail, it will be a sample of some response bias.

Then the next steps would be considering the weighted stratification, the method to calculate the margin of error has strong robustness.

References

Picture 1: 2006, Open School BC, http://bclearningnetwork.com/LOR/media/SS11/COURSE_4034644_M/my_files/module1/section1/lesson2/topic2.htm

DATA CLEANING CODE from Rohan alexander, "gss_cleaning.R", 7 Oct. 2020. tellingstorieswithdata.com

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
https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DUS88V

Technology, A. (2020, April). Data Centre. Retrieved December 10, 2020, from http://dc.chass.utoronto.ca/myaccess.html

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.