

Justin Trudeau would still win 2019 Canadian Federal Election even if “everyone” had voted*

Wen Wang

21 December, 2020

1. Introduction

The voting process is crucial since it would determine the Prime Minister/President of a country. However, this is not true that everyone eligible to vote had voted in the elections and this phenomenon would probably affect the final results. According to an article in *The Economist*, if “everyone” had voted in 2016 US presidential election, Hillary Clinton would probably be president.[*The Economist*]. Based on the information on Statistics Canada, just over three-quarters (77%) of Canadians reported voting in the 2019 federal election [Statistics Canada], so it could be inferred that Andrew Scheer would probably win the election if “everyone” had voted.

So, how to make “everyone” vote? It is time-consuming and cost-consuming and almost impossible to do a survey among the target population, all Canadians citizens who are above 18 in this case since not everyone would like to participate in the survey. One statistic way to solve this is through multilevel regression and poststratification (MRP). The central idea of MRP is to partition the data into thousands of demographic cells, estimate voter intent at the cell level using a multilevel regression model, and finally aggregate the cell-level estimates in accordance with the target population’s demographic composition.[et al., 2015]. In this report, I will use MRP to identify whether Andrew Scheer or Justin Trudeau would win the 2019 Canadian federal election if “everyone” had voted.

The proceed is as follows. In the Methodology section (Section 2), two data sets (survey data and census data) will be introduced to investigate how MRP could be used to predict whether Andrew Scheer or Justin Trudeau would win the 2019 Canadian federal election if “everyone” had voted. Results of the MRP analysis are provided in the Results section (Section 3), and summary and conclusion of the MRP analysis along with Weaknesses and Next Steps are presented in Conclusion section (Section 4).

2. Methodology

2.1 Data

The survey data is from 2019 Canadian Election Study, “ces2019_web” [Stephenson et al., 2020]. The original data is

The census data is from General social survey on Family (cycle 31), 2017 [Faculty of Arts & Sciences, b]. The GSS program, established in 1985, conducts telephone surveys across the ten provinces. The GSS is recognized for its regular collection of cross-sectional data that allows for trend analysis, and its capacity to test and develop new concepts that address current or emerging issues. For the census data survey, the target population is all people in Canada who are above 15 years old, excluding those from Yukon, Northwest Territories, and Nunavut. Full-time residents of institutions are also excluded. The frame population is

*Code and data are available at: https://github.com/wangw218/STA304_FinalProject

people who could be contacted by telephones. The sampled population is people who took the survey by telephones. Each record in the survey frame was assigned to a stratum within its province. A simple random sample without replacement of records was next performed in each stratum. There are 20,602 respondents and 461 variables in the census data.[Faculty of Arts & Sciences, a]

The two plots: Figure ?? and Figure ?? shows the raw data information about the provinces of the Survey data and Census Data. From these two plots, we could see Ontario has the largest number of samples, following by Quebec in both of these two datasets. Alberta and British Columbia have the larger number of respondents compared with left. One thing is note is that in our census data survey, people from Yukon, Northwest Territories, and Nunavut are excluded. Therefore, there are no information about these three provinces in Figure ?. The number of respondents from three three territories are all approximately 0 from the survey data according to the Figure ?? so these three territories would not have a great influence on the result.

4. Discussion

4.1 Summary

4.2 Weaknesses

4.3 Next Steps

Appendix

In this report, we used R [R Core Team, 2020], Rstudio [RStudio Team, 2020] and R packages: tidyverse [Wickham et al., 2019], lme4 [Bates et al., 2015], knitr [Xie, 2019] [Xie, 2015] [Xie, 2014], cesR [Hodgetts and Alexander, 2020] and labelled [Larmarange, 2020]

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

Stephenson, Laura, Allison Harrel, Daniel Rubenson and Peter Loewen. Forthcoming. ‘Measuring Preferences and Behaviour in the 2019 Canadian Election Study,’ Canadian Journal of Political Science

References

- Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015. doi: 10.18637/jss.v067.i01.
- W. Wang et al. Forecasting elections with non-representative polls. *International Journal of Forecasting*, 31(3):980–991, 2015. doi: <https://doi.org/10.1016/j.ijforecast.2014.06.001>.
- University of Toronto Faculty of Arts & Sciences. 2017 general social survey (gss): Families cycle 31 user’s guide, a. URL https://sda-artsci-utoronto-ca.myaccess.library.utoronto.ca/sdaweb/dli2/gss/gss31/gss31/more_doc/GSS31_User_Guide.pdf.
- University of Toronto Faculty of Arts & Sciences. Canadian general social surveys (gss), b. URL <https://sda-artsci-utoronto-ca.myaccess.library.utoronto.ca/sdaweb/html/gss.htm>.
- Paul A. Hodgetts and Rohan Alexander. *cesR: Access the CES Datasets a Little Easier.*, 2020. R package version 0.1.0.
- Joseph Larmarange. *labelled: Manipulating Labelled Data*, 2020. URL <https://CRAN.R-project.org/package=labelled>. R package version 2.7.0.

- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020. URL <https://www.R-project.org/>.
- RStudio Team. *RStudio: Integrated Development Environment for R*. RStudio, PBC, Boston, MA, 2020. URL <http://www.rstudio.com/>.
- Statistics Canada. Reasons for not voting in the federal election, october 21, 2019. URL <https://www150.statcan.gc.ca/n1/daily-quotidien/200226/dq200226b-eng.htm>.
- Laura B Stephenson, Allison Harell, Daniel Rubenson, and Peter John Loewen. 2019 Canadian Election Study - Online Survey, 2020. URL <https://doi.org/10.7910/DVN/DUS88V>.
- The Economist. If everyone had voted, hillary clinton would probably be president. URL <https://www.economist.com/graphic-detail/2019/07/06/if-everyone-had-voted-hillary-clinton-would-probably-be-president>.
- Hadley Wickham, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Grolmund, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Lin Pedersen, Evan Miller, Stephan Milton Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Paige Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo, and Hiroaki Yutani. Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686, 2019. doi: 10.21105/joss.01686.
- Yihui Xie. knitr: A comprehensive tool for reproducible research in R. In Victoria Stodden, Friedrich Leisch, and Roger D. Peng, editors, *Implementing Reproducible Computational Research*. Chapman and Hall/CRC, 2014. URL <http://www.crcpress.com/product/isbn/9781466561595>. ISBN 978-1466561595.
- Yihui Xie. *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition, 2015. URL <https://yihui.name/knitr/>. ISBN 978-1498716963.
- Yihui Xie. *knitr: A General-Purpose Package for Dynamic Report Generation in R*, 2019. URL <https://yihui.name/knitr/>. R package version 1.23.