

Informative Title Name

STA304 - Assignment 2

Eric Zhu, Yumeng Wang, Rafee Rahman

November 24, 2022

Introduction

The Canadian Constitution requires federal elections to be held at least every five years. By convention, Canada has been holding federal elections every four years. They are generally held on the third Monday of October in the fourth year after the last federal election. In some cases, elections may break this pattern as the governor general can dissolve parliament and call an election [1].

Canada is divided into 338 constituencies, also known as ridings [1]. Canada runs on a political party system where each party has one leader, the candidate running for Prime Minister, and representatives of that party attempting to win over a riding. On election day, registered voters will vote in the polling divisions within their riding, and select one member of parliament representing the party of their choice. The leader of the Canadian party that wins over the majority of ridings (170 of the 338) will become the Prime Minister.

With the next Canadian federal election taking place in 2025, **the goal of our report is to predict the overall popular vote using a regression model with post-stratification**. Post-stratification is a way to adjust the weighted totals within mutually exclusive cells so that they equal the known population totals [2]. This helps account for differences between some survey populations, and the true population. By post-stratifying, we will obtain better results and make better inferences about the true population.

The most popular vote is determined by the political party that has the highest number of votes, and this outcome will be the party that is most likely to win the election. This might not always be the case, however, because ridings vary in population. The party with the highest number of votes may not always win because another party might have won more ridings with narrow margins [3].

Using logistic regression models and survey data about the party choice of each voter, we will use predictor variables including province, sex, visible minority, income, and age to make our predictions about whether a party is voted for or not. In the post-stratification model, we will weigh the predicted outcome for each cell that holds the variables of interest, by the proportion of the population. This will help adjust the data correctly to make correct inferences for the actual population.

Based on the percentage of popular votes being slightly higher for the Conservatives than the Liberal party during the 2021 and 2019 federal elections [4], **we will hypothesize that the overall popular vote will remain the same and go to the Conservative Party**.

Data

<Type here a paragraph introducing the data, its context and as much info about the data collection process that you know.>

<Type here a summary of the cleaning process (**only add in stuff beyond my original gss_cleaning.R code**). You only need to describe additional cleaning that you and your group did.>]

<Here is a resource for grabbing the CES2019 data: <https://awstringer1.github.io/sta238-book/section-short-tutorial-on-pulling-data-for-assignment-1.html#section-canadian-election-study>>

<Remember, you may want to use multiple datasets here, if you do end up using multiple data sets, or merging the data, be sure to describe this in the cleaning process and be sure to discuss important aspects of all the data that you used.>

<Include a description of the important variables.>

<Include a description of the numerical summaries. Remember you can use `r` to use inline R code.>

Use this to create some plots. Should probably describe both the sample and population.

<Include a clear description of the plot(s). I would recommend one paragraph for each plot.>

Methods

<Include some text introducing the methodology, maybe restating the problem/goal of this analysis.>

Model Specifics

<I will (incorrectly) be using a linear regression model to model the proportion of voters who will vote for Donald Trump. This is a naive model. I will only be using age, which is recorded as a numeric variable, to model the probability of voting for Donald Trump. The simple linear regression model I am using is:>

$$y = \beta_0 + \beta_1 x_{age} + \epsilon$$

<Where y represents the β_0 represents....>

Post-Stratification

<In order to estimate the proportion of voters....>

<To put math/LaTeX inline just use one set of dollar signs. Example: \hat{y}^{PS} >

include.your.mathematical.model.here.if.you.have.some.math.to.show

All analysis for this report was programmed using **R version 4.0.2**.

Results

<Here you present your results. You may want to put them into a well formatted table. Be sure that there is some text describing the results.>

<Note: Alternatively you can use the `knitr::kable` function to create a well formatted table from your code. See here: <https://rmarkdown.rstudio.com/lesson-7.html>.>

<Remember you can use `r` to use inline R code.>

<Include an explanation/interpretation of the visualizations. Make sure to comment on the appropriateness of the assumptions/results.>

Conclusions

<Here you should give a summary of the Hypotheses, Methods and Results>

<Highlight Key Results.>

<Talk about big picture.>

<Comment on any Weaknesses.>

<End with a concluding paragraph to wrap up the report.>

Bibliography

1. Golemund, G. (2014, July 16) *Introduction to R Markdown*. RStudio. https://rmarkdown.rstudio.com/articles_intro.html. (Last Accessed: January 15, 2021)
2. Dekking, F. M., et al. (2005) *A Modern Introduction to Probability and Statistics: Understanding why and how*. Springer Science & Business Media.
3. Allaire, J.J., et. el. *References: Introduction to R Markdown*. RStudio. <https://rmarkdown.rstudio.com/docs/>. (Last Accessed: January 15, 2021)

(Will cite properly at the end)

1 <https://www.thecanadianencyclopedia.ca/en/article/electoral-systems> 2 <https://www.surveypractice.org/article/2809-post-stratification-or-non-response-adjustment> 3 <https://www.cicnews.com/2021/08/how-canadas-electoral-system-works-0819016.html#gs.j2j5wh> 4 <https://www.sfu.ca/~aheard/elections/1867-present.html>