Analyzing Provincial Electoral Support of the Liberal Party: A Canadian Election Study

STA304 - Winter 2025 - Assignment 2

Group 120: Mohammad Danish Malik, Abdul Ahad Qureshi, Syed Hasan

1 Introduction

Public opinion surveys are crucial for understanding political voter behavior and predicting electoral outcomes. The 2019 Canadian Federal Election Study (CES) contains one such important Web Survey that captures distinct voting patterns and demographic characteristics across Canada. The Web Survey provides researchers with a rich resource to ascertain how different groups support political parties and to predict the strength of the support for any political party.

This report investigates the electoral support for the Liberal Party, adopting a stratified random sampling approach from the 2019 CES Web Survey. In this analysis, we explore how residence in a particular province of Canada, along with the age of the voter, influences the odds of voting Liberal. We aim to uncover the extent to which province stratification affects Liberal Party support.

We aim to first calculate and provide a robust estimate of the Liberal Party's expected vote share and compute a 95% confidence interval around this estimate. We will then create a logistic regression model to predict the log odds of voting Liberal based on a voter's province of residence and age. Through this dual approach, we aim to address the research question: ***how much does a voter's province of residence in Canada and age influence their likelihood of voting for the Liberal Party?***

2 Data

Our study used the Web survey data set of the Canadian 2019 election study. Our primary goal was to analyze the relationship between age, province and the likelihood of voting Liberal and these were the variables of interest. In the web data set these translate to cps19_yob (age), cps19_province (province) and cps19_votechoice (vote). Age was calculated from the

respondents' year of birth, province was treated as a categorical factor and used as the stratification variable and vote choice was transformed into a binary variable indicating whether the respondent voted for the Liberal Party. To ensure the dataset was clean and structured for analysis, we performed data cleaning and variable transformations.

One of the key steps in data cleaning was the mapping of province codes from cps19_province to their corresponding province names.This transformation ensured that provinces were treated as categorical variables in the analysis, making the results more interpretable. We also converted province into a factor variable to ensure it was handled correctly in the logistic regression model.

The next step was handling voting data by transforming the cps19_votechoice variable into a new variable called vote with all numeric codes mapped to their respective parties. A binary variable called is_liberal was also created. This took the values 0 if the respondent did not vote for the liberal party and the value 1 if they did vote for liberals. This transformation allowed us to model the probability of voting Liberal as a function of age and province. We also filtered out any missing values of data which would skew our results. Converted year of birth (cps19_yob) to age. After this we created a new dataset called voter_data which included variables: age, province, is_liberal, vote, cps19_votechoice, cps19_province. After these cleaning steps, the dataset was ready for analysis. Finally, to visualize the distribution of the stratification variable we created **Figure 3** (see Results section). This displays the number of survey respondents by province, confirming adequate representation across regions and validating the stratified method of the data collection.

3 Methods

We calculate a confidence interval (CI) to measure the reliability of our estimated proportion of Liberal Party voters. The 95% CI provides a range within which the true proportion is likely to fall, accounting for sampling variability. This is the formula used for the confidence interval:

$$CI = \hat{p} \pm 1.96 \times SE(\hat{p})$$

where:

- $\hat{p}$= weighted estimate of the proportion
- $SE(\hat{p})$= standard error accounting for stratification

The standard error (SE) measures the variability of our estimated proportion of Liberal Party voters due to sampling randomness. A smaller SE indicates a more precise estimate while a larger SE suggests greater uncertainty in our results.

$$SE(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

where:

- $\hat{p}$= proportion of Liberal voters
- n= total number of respondents

We used a logistic regression model understand the relationship between age, province, and the probability of voting for the Liberal Party. The model below estimates how these factors influence voting behavior by predicting the log-odds of supporting the Liberal Party. The probability of voting Liberal was modeled as a function of province and age. Our model allows comparisons of different provinces while controlling for age.

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 \times \text{age} + \sum_k \beta_k \times \text{province}_k$$

where:

- π= probability of voting for the Liberal Party
- β_0= intercept
- β_1= coefficient for age
- β_k= coefficients for each province

Our logistic regression model allows us to quantify the impact of age and regional differences on voting preferences. The intercept, β_0, represents the log odds of voting for the Liberal Party in the reference category (Alberta). β_1 (age coefficient) measures how age affects the probability of voting Liberal. If β_1 is negative, it implies that older voters are less likely to vote for the Liberal Party. β_k captures the effect of residing in a specific province on the likelihood of voting Liberal. The reference category for province is Alberta which means that all other provinces are compared against Alberta. A positive β_k for a province suggests higher support for the Liberal Party relative to Alberta, while a negative β_k suggests lower support. The magnitude of each β_k determines the strength of the effect.

In addition to the regression model, we created three key visualizations to support our findings. First, a histogram of respondent ages provided insight into the age distribution within the survey. Second, a bar plot of Liberal vote proportions by province illustrated regional differences in political preferences. Finally, a bar plot of respondent counts by province helped assess how well-represented each region was in the data set.

4 Results

The estimated proportion of respondents who reported voting for the Liberal Party was:

$$\hat{p} = 0.336$$

The 95% confidence interval for this estimate ranges from 0.330 to 0.341. This is a small confidence interval that suggests a high level of precision, indicating that between 33.0% and 34.1% of the population is estimated to have voted for the Liberal Party. The results show that at least 1/3 of voters supported the liberal party.

Table 1: Confidence Interval for Proportion of Votes for the Liberal Party

Estimate	Standard Error	**95% CI Lower**	**95% CI Upper**
0.336	0.0028	0.330	0.341

Below is the estimate regression model:

$$\begin{aligned}\log\left(\frac{\pi}{1-\pi}\right) = & -1.2697 + 0.7146 \times \text{British Columbia} + 0.6611 \times \text{Manitoba} \\ & + 1.0657 \times \text{New Brunswick} + 1.4889 \times \text{Newfoundland} \\ & + 1.3288 \times \text{Northwest Territories} + 1.3545 \times \text{Nova Scotia} \\ & + 1.2554 \times \text{Nunavut} + 1.0934 \times \text{Ontario} \\ & + 1.1902 \times \text{Prince Edward Island} + 0.9446 \times \text{Quebec} \\ & - 0.2685 \times \text{Saskatchewan} + 0.5078 \times \text{Yukon} \\ & - 0.0054 \times \text{age}\end{aligned}$$

Table 2: Logistic Regression Model Estimates

term	estimate	std.error	statistic	p.value
(Intercept)	-1.2697	0.0608	-20.8964	0.0000
provinceBritish Columbia	0.7146	0.0609	11.7275	0.0000
provinceManitoba	0.6611	0.0789	8.3818	0.0000
provinceNew Brunswick	1.0657	0.0970	10.9817	0.0000
provinceNewfoundland & Labrador	1.4890	0.1061	14.0285	0.0000
provinceNorthwest Territories	1.3288	0.4859	2.7349	0.0062
provinceNova Scotia	1.3545	0.0883	15.3445	0.0000
provinceNunavut	1.2554	0.4785	2.6236	0.0087

term	estimate	std.error	statistic	p.value
provinceOntario	1.0934	0.0506	21.6087	0.0000
provincePrince Edward Island	1.1902	0.2073	5.7426	0.0000
provinceQuebec	0.9446	0.0540	17.4802	0.0000
provinceSaskatchewan	-0.2685	0.1037	-2.5907	0.0096
provinceYukon	0.5078	0.4512	1.1254	0.2604
age	-0.0054	0.0008	-6.8670	0.0000

Interpretation of Model Estimates

The regression model shows that both province and age are significant predictors of the likelihood of voting liberal. The intercept of -1.2697 shows the log odds of voting liberal for a respondent from Alberta at age 0. The other provinces show statistically significant and positive coefficients relative to Alberta, this suggests higher support for the Liberal Party. Newfoundland (B=1.4889) and Nova Scotia (B=1.3545) show the strongest positive associations. Similarly, Ontario (B=1.0934) and Quebec (B=0.9446) demonstrate substantially greater support for the Liberal Party relative to Alberta.Saskatchewan (B= -0.2685) shows a statistically significant negative coefficient, indicating lower odds of voting Liberal compared to Alberta. Yukon has a positive coefficient but does not show statistical significance with a high p value (p=0.25). Age is also an important predictor. With a coefficient of B = -0.0054, it implied that older respondents are less likely to support the liberal party. Each additional year of age results in a decrease of voting liberal.

With our model, there are also a few figures which have been created. **Figure 1** shows the age distribution of respondents. The data shows a uniform spread across age groups, with the highest representation among respondents aged 35 - 65. This shows a strong middle aged population in the sample. **Figure 2** displays the proportion of Liberal voters by province. We see clear regional differences in support. Newfoundland & Labrador, Nova Scotia, Ontario show the highest proportions of Liberal support going over 40%. On the other hand, Saskatchewan and Alberta have the lowest Liberal support proportions. **Figure 3** shows the number of respondents per province. Ontario had the largest number of respondents, followed by Quebec and Alberta. The smaller provinces and territories had fewer respondents, consistent with their population sizes.

Figure 1: Age Distribution of Respondents

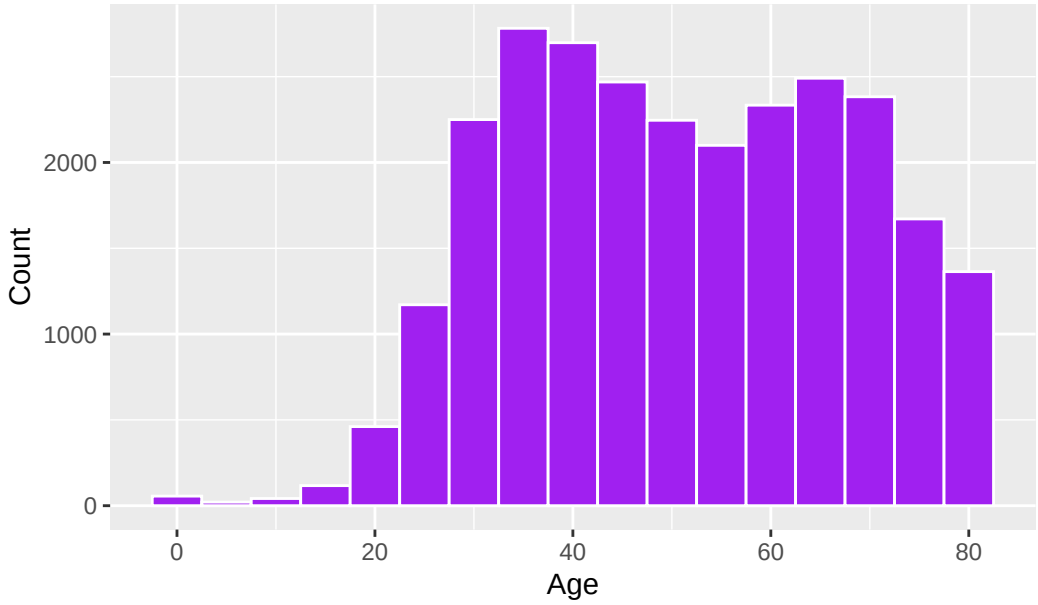


Figure 2: Proportion of Liberal Voters by Province

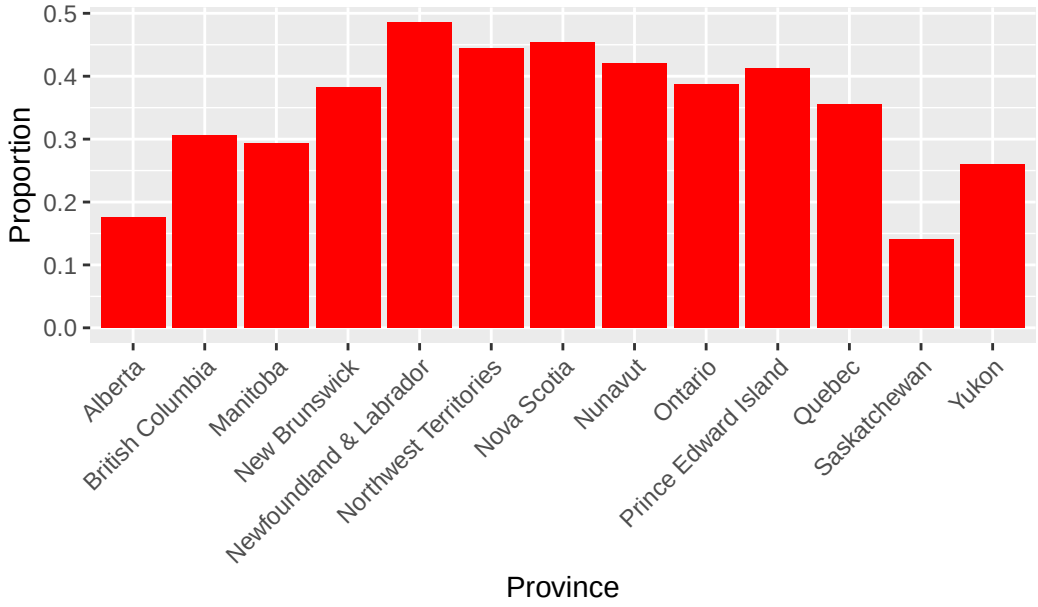
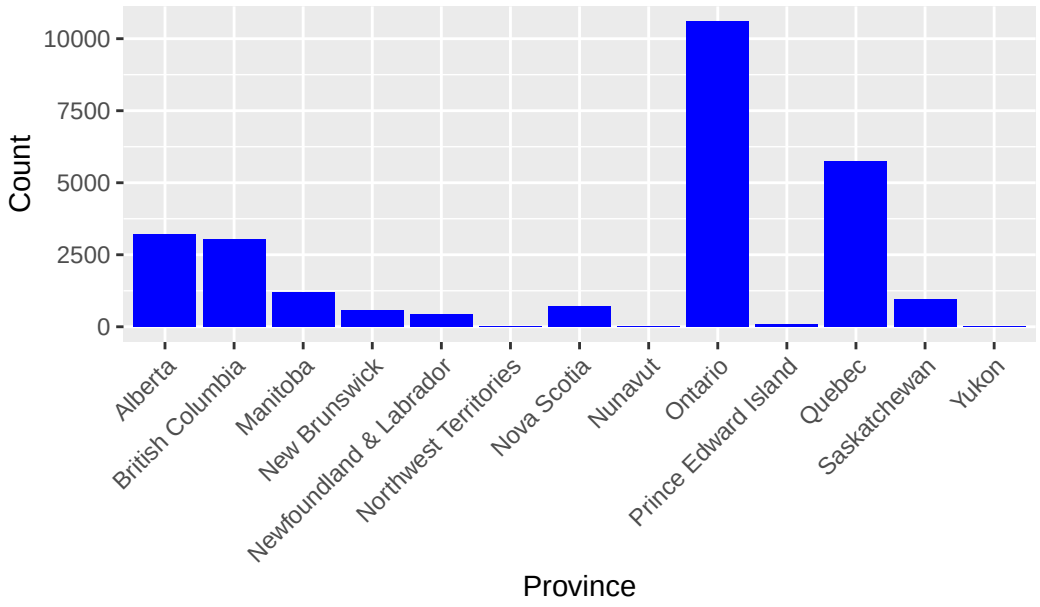


Figure 3: Number of Respondents per Province



Key Takeaways

There are a few key takeaways from our results. One third of Canadian respondents reported voting for the liberal party, with a high level of precision in the estimate. Newfoundland & Labrador and Nova Scotia show strong liberal support when compared relative to Alberta. Age negatively correlates to liberal support where older individuals are less likely to vote.

5 Discussion

Overview of Key Findings

This report aimed to examine the influence of province and age on the likelihood of voting for the Liberal Party. The key findings from the results section show significant regional and age-based variations in voting preferences. A voter's province of residence plays a significant role in shaping support for the Liberal Party, with higher proportions of Liberal voters in regions like Nunavut, Newfoundland & Labrador, and Nova Scotia, while Alberta and Saskatchewan show much lower levels of support for the Liberal Party. Age also significantly impacts voting behavior, with older individuals being less likely to vote Liberal, thus highlighting the importance of age demographics in electoral engagement.

The results underscore the need to consider both geographic and age-related factors when analyzing political preferences in Canada. The variation in Liberal Party support across provinces suggests the importance of targeting campaigns based on regional dynamics. Additionally, the negative association between age and the likelihood of voting Liberal aligns with broader trends

observed in other democracies, where older voters tend to support more conservative parties while younger voters lean toward more liberal or progressive options.

Limitations of the Analysis

While the study provides valuable insights into the relationship between province, age, and voting preferences, several limitations must be considered. The self-reported nature of survey responses introduces potential biases. While self-reporting can provide valuable insights into individuals' political preferences, it also leaves room for inaccuracies. Respondents may misinterpret questions, exaggerate their political preferences, or overstate their intention to vote. Moreover, the survey methodology introduces its own set of biases. Web surveys, though anonymous, suffer from selection bias, as they tend to underrepresent individuals with limited internet access while attracting more politically engaged participants (Bethlehem, 2010). Multitasking and distractions during web surveys can also affect the quality of the data (Callegaro et al., 2015). However, web respondents may provide less socially desirable answers due to the anonymity offered, although the absence of interviewer guidance can lead to misinterpretations (Booth-Kewley et al., 2007).

Future Recommendations for Research

Future research could benefit from addressing these limitations by incorporating additional demographic variables such as income, education, and ethnicity to provide a more comprehensive understanding of voting behavior and political preferences, particularly when examining their intersection with regional and age-based differences. Further studies should also explore the impact of survey methodologies, specifically comparing phone and web surveys, as phone surveys, while offering the benefit of live interaction, can introduce interviewer bias (Gideon, 2012), while web surveys may suffer from selection biases. A deeper understanding of these biases would help refine survey methodologies for more accurate and representative political predictions. Additionally, incorporating longitudinal data would enable researchers to track shifts in voter preferences over time, offering valuable insights into how age-related changes in political engagement influence party support, especially as younger voters tend to lean more toward liberal or progressive parties, while older individuals are more likely to support conservative ideologies (Friedman & Schultz, 2024). Understanding these trends would be crucial for political parties aiming to mobilize younger voters and secure future electoral support.

Conclusion

This analysis highlights how province and age influence voting behavior in the 2019 Canadian Federal Election, emphasizing the need for targeted political campaigns based on these factors. However, limitations in survey methodology, sample biases, and omitted demographic variables like income and education may affect the generalizability of the results. The use of web surveys introduces selection bias and potential distractions, which could influence the data quality. Future research should address these issues by incorporating additional demographic variables, refining survey methodologies, and exploring biases in different survey modes, such as phone versus web surveys. Including longitudinal data would also provide insights into how age-related shifts in political engagement affect party support. These improvements will

enhance the accuracy of predictions and help political parties better engage younger voters in future elections.

6 Generative AI Statement

We used generative AI tools to assist with writing structure, fixing code errors, and refining explanations. We ensured that the final report was our own work and met the assignment's requirements by reviewing and modifying the AI-generated content. The statistical analysis, coding, and interpretation were conducted independently with AI only serving as a a supplementary tool to enhance clarity. Any AI-generated text was rewritten in our own words to maintain academic integrity, and all analysis was cross-checked against course materials and data set documentation to ensure accuracy.

7 Ethics Statement

Ensuring the reproducibility of our analysis was a key ethical consideration in our project. We achieved this by documenting our data processing steps, statistical methodology, and analysis workflow. The code used for data cleaning, transformations, and statistical modeling is well-structured and includes comments to enhance clarity. Additionally, we employed functions such as "set.seed()" to ensure that any random sampling or simulations yield consistent results which allows for the study to be replicated.

The 2019 Canadian Federal Election Study (CES) data set used in this analysis is publicly available which eliminates any concerns regarding confidentiality breaches. The CES data set has been anonymized, meaning that no personally identifiable information (PII) is included, and individual respondents cannot be traced. This ensures that participants' rights to privacy and confidentiality are maintained. Since the data set is anonymized and de-identified, there are no direct risks to participants and Research Ethics Board (REB) approval is not required for making this report publicly available. However, ethical considerations concerning privacy, misrepresentation, and potential biases in survey research would still remain. Our study mitigated these risks by employing stratified random sampling assumptions and rigorous statistical methods to estimate confidence intervals and model voting behavior. The interpretation of our results includes a discussion of limitations to ensure that findings are not overstated or taken out of context. Furthermore, we have refrained from reporting findings in a way that could unfairly stereotype or misrepresent demographic groups, ensuring that our interpretations remain responsible and within ethical guidelines.

We must consider the impacts and rights of different stakeholders affected by this analysis. From an impact perspective, ensuring transparency in our methodology enhances the credibility of our findings and contributes to a broader understanding of voting behavior in Canada.By clearly presenting our methods and discussing the limitations of our study, we aim to prevent

misinformation and promote responsible use of survey data. From a rights-based perspective, participants in the original survey have the right to privacy and confidentiality. The CES data set protects these rights by removing identifiable information, but ethical obligations still exist in how findings are presented. Misrepresentation of results could lead to stereotyping or misleading narratives about certain demographic groups, which could have real-world consequences, such as shaping public opinion or influencing policy in unintended ways. To address this, we have carefully interpreted our results, avoiding over generalization and acknowledging the limitations of survey methodologies. We aimed to follow the best practices in data analysis, transparency, and ethical reporting to uphold the integrity of our research and contribute to the broader goal of responsible statistical inquiry.

8 Bibliography

1. Bethlehem, J. (2010). Selection bias in web surveys. International Statistical Review, 78(2), 161–188. https://doi.org/10.1111/j.1751-5823.2010.00112.x
2. Booth-Kewley, S., Larson, G. E., & Miyoshi, D. K. (2007). Social desirability effects on computerized and paper-and-pencil questionnaires. Computers in Human Behavior, 23(1), 463–477. https://doi.org/10.1016/j.chb.2004.10.020
3. Callegaro, M., Manfreda, K. L., & Vehovar, V. (2015). Survey research and web surveys. SAGE Publications Ltd. https://doi.org/10.4135/9781529799651
4. Friedman, S., & Schultz, D. (2024). Generational politics in the United States: From the silents to Gen Z and beyond(1st ed.). University of Michigan Press. https://doi.org/10.3998/mpub.11621506
5. Gideon, L. (Ed.). (2012). Handbook of survey methodology in social sciences (1st ed.). Springer. https://doi.org/10.1007/978-1-4614-3876-2
6. OpenAI. (2025). ChatGPT (February 6, 2025 version) [Large language model]. OpenAI. https://openai.com/chatgpt