Are Women More Likely To Vote Liberal?

STA304 - Winter 2025 - Assignment 2

GROUP NUMBER: 77

1 Introduction

Understanding voting behavior is essential for assessing democratic engagement and representation. Prior research has highlighted the importance of sociodemographic factors, including gender, in shaping electoral preferences and participation (Gidengil et al., 2006). In Canada, gender gaps in political attitudes and party preferences have been observed, with studies suggesting that women are more likely to support left-leaning parties compared to men (Frederick et al., 2009). However, the extent to which gender influences voter preferences in specific elections, such as the 2019 Canadian Federal Election, remains an open question.

This study examines the relationship between gender and voting intentions in the 2019 Canadian Federal Election using data from the Canadian Federal Election Study (CES) collected through a phone survey, in particular it looks at whether women were more prone than men to vote for the liberal party. The CES employed a stratified random sampling approach, ensuring representation across the gender groups. However, despite a nearly equal gender distribution in the Canadian population, female respondents were underrepresented in the survey sample, comprising only 41.7% of the 2,769 respondents. The primary outcome of interest in this study is voting intention for the Liberal Party, analyzed in relation to gender

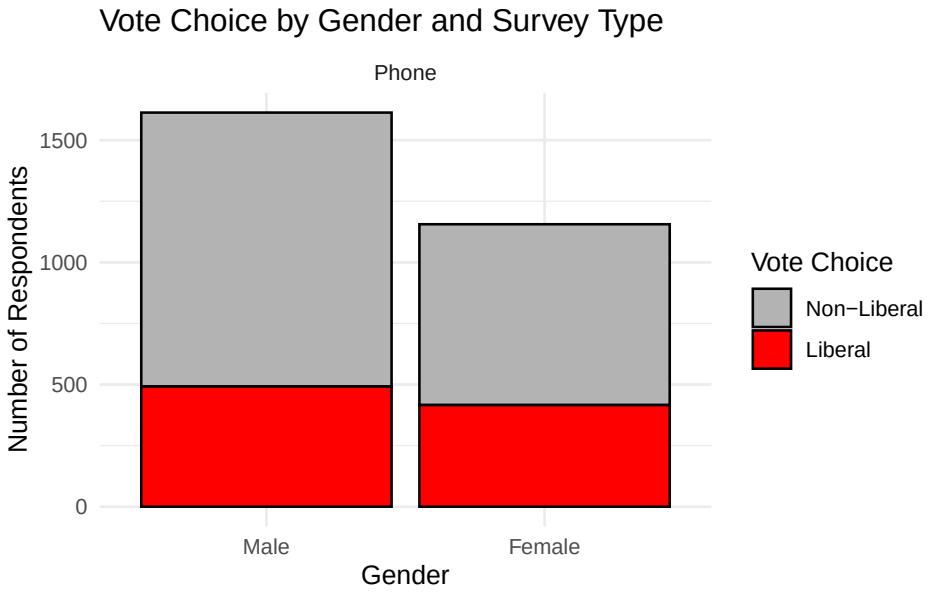
The results of this study contribute to the broader literature on gender and voting behavior in Canada, offering insights into the electoral dynamics of the 2019 federal election.

2 Data

The analysis draws from the 2019 Canadian Federal Election Study (CES), which was collected using stratitifed random sampling by gender. Gender contained three levels, male, female and other. The total male population and total female population were retrieved from the 2021 Census of Population Statistics Canada. In 2021, 50.7% of the Canadian population were women (18.77 mil-

lion out of 37.6 million). For the CES, data was obtained via a phone survey where 2,769 responses were obtained with only 41.7% of the respondents being female. The participants were asked a series of questions regarding their sociodemographic background and their intent to vote in the upcoming election, including interest in the election, likelihood to vote, and intended vote choice.

In this study, the data cleaning involved three main steps. First, we dichotomized the primary outcome variable, voting intention for the Liberal party. Second, entries with missing or invalid responses for intended party or gender were excluded. Lastly, only male and female responders were considered given the small number of people in the other gender categories. The small sample size of the other category could limit meaningful statistical analysis and may have led to unreliable estimates. Given this, we chose to exclude this category to ensure more stable and interpretable results. We acknowledge that this decision reduced the inclusivity of our analysis and may overlook important differences in experiences. Future research with a larger and more representative sample should aim to better capture gender diversity.



Note The bar plot displays vote choice by gender, with the left column representing male respondents. The y-axis shows the number of respondents with that gender, while the colors indicate vote choice: red for Liberal and gray for Non-Liberal. Males have a higher overall response count compared to females, but the proportion of Liberal to Non-Liberal voters seems slightly higher for females, given that the overall number of liberal voters is similar across both genders.

3 Methods

This study investigates whether gender influences the likelihood of voting for the Liberal Party, using data from the 2019 Canadian Federal Election Study (CES), which was collected through stratified random sampling. Since stratified sampling ensures representation across key subgroups, all statistical estimates—including the proportion of Liberal voters and the logistic regression model predicting voting likelihood—account for survey weighting and finite population correction (FPC) (Lohr, 2019).

To estimate the proportion of Liberal voters, we use a weighted mean across strata. The estimated proportion is calculated as:

$$\hat{p}_{st} = \sum_{h=1}^{H} W_h \hat{p}_h$$

where H represents the number of strata (e.g., provinces or education levels), $W_h = N_h/N$ is the stratum weight based on its share of the total population, and $\hat{p}_h$ is the proportion of Liberal voters within each stratum. The 95% confidence interval (CI) is given by:

$$CI = \hat{p}_{st} \pm z_{\alpha/2} \sqrt{\sum_{h=1}^{h} W_h^2 \left(1 - \frac{n_h}{N_h}\right) \left(\frac{\hat{p}_h(1-\hat{p}_h)}{n_h}\right)}$$

where n_h is the sample size within each stratum, N_h is the population size in that stratum, and $(1 - n_h/N_h)$ is the finite population correction (FPC), which accounts for cases where the sample represents a substantial fraction of the population (Ziegel et al.,2000). Without this adjustment, confidence intervals could be overly wide, leading to inflated uncertainty.

To analyze the relationship between gender and voting preference, we fit a survey-weighted logistic regression model, which adjusted for stratification effects and unequal selection probabilities. Since the dependent variable (voting for the Liberal Party) was binary (1 = Yes, 0 = No), logistic regression was the appropriate modeling choice (Lumley, 2010). The model is specified as:

$$\log\left(\frac{P(VoteLiberal = 1)}{1 - P(VoteLiberal = 1)}\right) = \beta_0 + \beta_1 \times \text{Gender}_{\text{male}} + \beta_2 \times \text{Age}$$

Here, β_1 represents the effect of gender. If the corresponding odds ratio e^{β_1} is greater than 1, it suggests that women are more likely to vote for the Liberal Party compared to men. If the odds ratio is less than 1, it suggests women are less likely to vote Liberal compared to men. Similarly, β_2 captures the influence

of age. An odds ratio greater than 1 would indicate that an increase in age is associated with a higher likelihood of voting for the Liberal Party, while an odds ratio less than 1 suggests the opposite.

Since the dataset was collected using stratified sampling, failing to account for this in the regression model would lead to biased coefficient estimates and incorrect standard errors. We apply survey-weighted logistic regression (svyglm()) from the survey package in R, incorporating design weights and finite population correction (Lumley, 2010).

4 Results

Present a table showing the estimated proportion of votes for the selected party along with the 95% confidence interval, and include text describing this table and the key takeaways.

Table 1: Stratified Confidence Interval for Liberal Voters (Phone Survey).

Survey Group	Proportion Voting Liberal	95% Confidence Interval
Overall	**0.333**	**(0.315, 0.350)**
Male	**0.305**	**(0.283, 0.327)**
Female	**0.360**	**(0.332, 0.388)**

The proportion of respondents in the phone survey who reported voting Liberal was 0.333, with a 95% confidence interval of (0.315, 0.350), indicating that the true population proportion is likely within this range.

	Dependent variable		
Predictors	Odds Ratios	CI	p
(Intercept)	0.33	0.25 – 0.43	**<0.001**
sex [Male]	0.79	0.67 – 0.92	**0.004**
age	1.01	1.01 – 1.02	**<0.001**
Observations	2769		
R^2 / R^2 adjusted	0.008 / 0.007		

Note Logistic regression model for intent to vote liberal, using sex and age as predictors, were female is the baseline cateogry for sex

The results from the survey-weighted logistic regression indicate that gender and age are both significant predictors of voting for the Liberal Party. The odds ratio for gender (male) is 0.79% (95% CI: 0.67–0.92, p = 0.004), suggesting that men have 21% lower odds of voting Liberal compared to women, or equivalently, women have 27% higher odds of voting liberal than men. Since the odds ratio is lower than 1, this indicates a positive association between being female and

voting Liberal, due to the fact that female is the baseline category. Similarly, the odds ratio for age is 1.01 (95% CI: 1.01–1.02, $p < 0.001$), implying that older individuals are slightly more likely to vote Liberal. While the effect size for age is small, it is also statistically significant.

5 Discussion

The analysis reveals that women are more likely to vote liberal, supporting our initial hypothesis with statistically significant results. Findings indicate that women have 27% higher odds of voting Liberal compared to men.

Potential biases may arise from survey self-selection and response bias, as individuals with particular political ideals might be more inclined to participants. Additionally, some relevant variables that could affect voting behavior, such as socioeconomic status or regional differences that might have made more difficult the participation in the surveys. Survey errors, including misreporting or misunderstanding of questions, also present challenges in interpreting the findings.

A notable limitation of the analysis is the exclusion of the "Other" gender category, which resulted in the removal of an entire group from the analysis. This omission restricts our ability to fully understand voting behavior across all gender identities. We acknowledge that future research should not only incorporate a broader range of variables but also place a stronger emphasis on collecting and analyzing a greater number of observations from non-binary or other gender identifications. This approach will provide a more inclusive and comprehensive understanding of voting behavior and improve the generalizability of the conclusions drawn from the data.

6 Generative AI Statement

Here is where you can explain your usage of Generative AI tool(s). Be sure to reference any tools with inline citations.

In the completion of this assignment, generative AI (OpenAI, 2025) was used to assist in struc- turing the analysis and generating explanations related to the influence of survey methodology on political preferences. Specifically, AI tools helped draft and refine sections on the potential biases introduced by phone and web surveys, ensuring clarity and conciseness in presenting the comparative analysis. Additionally, AI was used to fix the formatting of visualizations and synthesize my initial draft of the paper before a subsequent re-write from my end, this was done to improve the overall coherence and readability of the final report. The use of gen- erative AI ensured a more streamlined process and provided valuable insights into improving the presentation of data.

7 Ethics Statement

The analysis was designed with reproducibility in mind, by thoroughly documenting the methods f the regression and data cleaning. The data cleaning process was explained step-by-step, clearly explaining which observations were not included in the analysis due to the small sample size, how is the data stratified, and which is our startification variable, ensuring that other researchers can precisely replicate our steps to verify the findings.

Since the CES 2019 dataset is publicly accessible and qualifies under U of T's Research Ethics Policy exemption criteria, our study does not require Research Ethics Board approval for the report to be publicly available. Moreover, the data have been anonymized and does not include direct identifiers, safeguarding the privacy of individual participants while still providing valuable insights into voting behavior.

8 Bibliography

1. Allaire, J.J., et. el. *References: Introduction to R Markdown*. RStudio. https://rmarkdown.rstudio.com/docs/. (Last Accessed: April 4, 1991)

2. Dekking, F. M., et al. (2005) *A Modern Introduction to Probability and Statistics: Understanding why and how.* Springer Science & Business Media.

3. Frederick, B. (2009). Are Female House Members Still More Liberal in a Polarized Era? The Conditional Nature of the Relationship Between Descriptive and Substantive Representation. Congress & the Presidency, 36(2), 181–202. https://doi.org/10.1080/07343460902948097

4. Gidengil, E., Everitt, J., & Fournier, P. (2006). Gender and Vote Choice in the 2006 Canadian Election. https://api.semanticscholar.org/CorpusID:140276235

5. Grolemund, G. (2014, July 16) *Introduction to R Markdown*. RStudio. https://rmarkdown.rstudio.com/articles_intro.html. (Last Accessed: April 4, 1991)

6. Vaughn, B. K. (2008). Data analysis using regression and multilevel/hierarchical models, by Gelman, A., & Hill, J [Review of *Data analysis using regression and multilevel/hierarchical models, by Gelman, A., & Hill, J*]. *Journal of Educational Measurement*, *45*(1), 94–97. Blackwell Publishing Inc. https://doi.org/10.1111/j.1745-3984.2007.00053_2.x

7. Ziegel, E. R., & Lohr, S. L. (2000). Sampling: Design and Analysis. *Technometrics*, *42*(2), 223-. https://doi.org/10.2307/1271491

8.

9 Appendix

Any additional notes/derivations that are supplementary to the report can be added in an appendix. This section will not be directly graded, but may be included for completion-sake.