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. Which is why in this study we will introduce the question: Are Women More Likely To Vote Liberal?

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

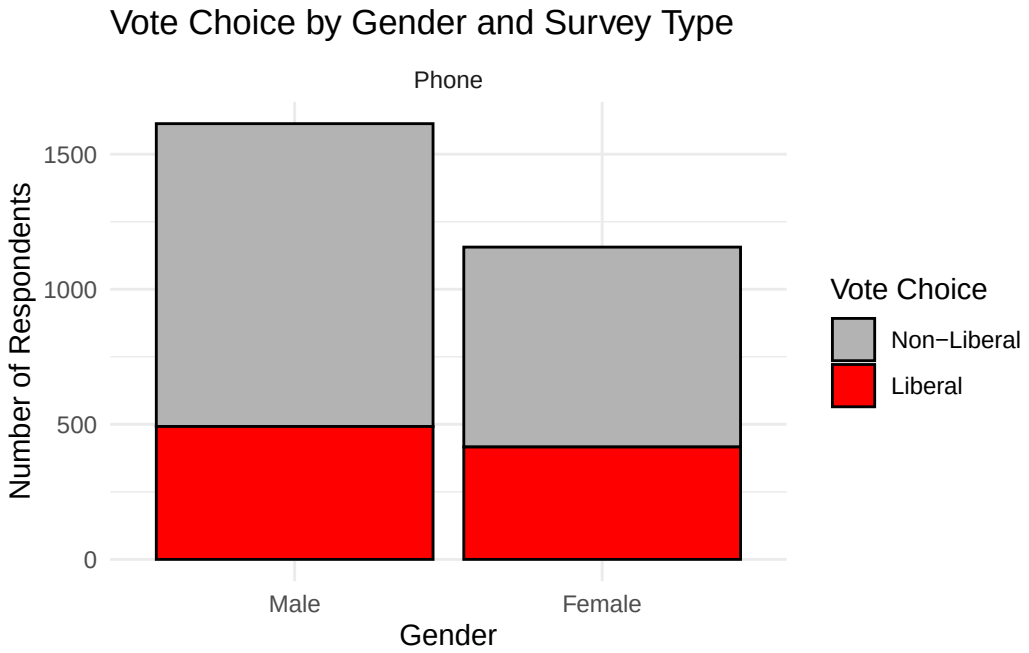
The results of this study contribute to the broader literature on gender and voting behavior in Canada, offering insights into the electoral dynamics of the 2019 federal election.

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 million 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., gender), $W_h = N_h/N$ 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, $z_{\frac{\alpha}{2}}$ is the Z-score corresponding to the desired confidence level (in this case 95%), 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:

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

We used logistic regression for this analysis because our dependent variable, voting for the Liberal Party, is binary (1 = Yes, 0 = No). This is because logistic regression models the log-odds of voting Liberal rather than predicting a direct probability, which allows to estimate

the likelihood of voting Liberal while ensuring that predicted probabilities stay within the valid range of 0 and 1.

β_1 represents the effect of being male on the log-odds of voting for the Liberal Party, meaning it captures how much the log-odds change when comparing males to the reference group, females. Since logistic regression models log-odds rather than direct probabilities, a negative β_1 means that the log-odds of voting Liberal decrease for males compared to females, suggesting that males are less likely to vote Liberal. To make this more intuitive, we exponentiated β_1 to obtain the odds ratio, e^{β_1}. If the odds ratio is less than 1, it confirms that males have lower odds of voting Liberal than females, while an odds ratio greater than 1 would indicate the opposite.

In the same way, β_2 represents the effect of age on voting Liberal, where a positive coefficient suggests that as individuals get older, their log-odds of voting Liberal increase. This means that older individuals are more likely to vote Liberal, and the odds ratio e^{β_2}quantifies how much the odds change with each additional year of age.

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

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

Survey Group	Proportion Voting Liberal	95% Confidence Interval
Overall	0.333	(0.315, 0.351)
Male	0.305	(0.283, 0.327)
Female	0.360	(0.332, 0.388)

Table 1 presents the estimated proportion of voters supporting the Liberal party, along with their 95% confidence intervals, calculated using a weighted mean across strata. The overall proportion of Liberal voters is 0.333, with a confidence interval of (0.315, 0.351), showing a stable estimate.

Looking at gender differences, females have a higher estimated proportion of Liberal voters (0.360, CI: 0.332, 0.388) compared to males (0.305, CI: 0.283, 0.327). Since these confidence intervals do not overlap much, it suggests that females are more likely to vote Liberal than males. This aligns with the logistic regression results, which also showed that males were

less likely to vote Liberal. The male and female total population values were retrieved from StatsCan, ensuring accurate weighting in the analysis. The "Other" category was omitted due to a small sample size, which made it difficult to compute stable survey weights and reliable confidence intervals. The finite population correction (FPC) was applied to adjust for sampling, ensuring the estimates accurately reflect the population.

Table 2: Survey-Weighted Logistic Regression Results

Variable	Estimate	Standard Error	95% CI Lower	95% CI Upper	p-value
(Intercept)	-1.120	0.140	-1.394	-0.845	0.000
sexMale	-0.239	0.082	-0.400	-0.079	0.004
age	0.011	0.002	0.006	0.015	0.000

Table 2 shows the logistic regression results, analyzing how gender and age influence the likelihood of voting Liberal. The results indicate that females are more likely to vote Liberal than males. The coefficient for males (β = -0.239, p = 0.004) is negative, meaning that being male decreases the log-odds of voting Liberal. Since the confidence interval (-0.400, -0.079) does not include zero, this effect is statistically significant. Age is positively associated with voting Liberal (β = 0.011, p < 0.001), meaning that as individuals get older, their log-odds of voting Liberal increase. The confidence interval (0.006, 0.015) confirms this significance. The intercept (β = -1.120, p < 0.001) represents the log-odds of voting Liberal for a female at age 0, suggesting that at the baseline, the probability of voting Liberal is low.

To better interpret these log-odds, I exponentiated the coefficients. Holding age constant, the odds of voting Liberal for males are exp(-0.239) 0.787 times the odds for females, confirming that males have a lower probability of voting Liberal. Similarly, for each additional year of age, the odds of voting Liberal increase by exp(0.011) 1.011 times, meaning that older individuals are slightly more likely to vote Liberal. Overall, Table 2 confirms that females are more likely to vote Liberal than males, and that age is positively associated with supporting the Liberal party.

5 Discussion

The objective of this study was to assess whether women were more likely to vote for the Liberal Party in the 2019 Canadian Federal Election. A key strength of the study is its use of stratified random sampling, the randomization of the data, and the documented procedures used and portrayed in the Methods section. In order to improve interpretability, we decided to portray the Odds Ratios, instead of log-odds, because we deemed them to be more intuitive when it comes to making conclusions in the study. The findings suggest potential mechanisms, such as gender differences in political ideology, that may explain the observed voting patterns. Most precisely, Women are more likely then men to vote for a Liberal party, the principal findings

indicate that women had 27% higher odds of voting Liberal compared to men, supporting our initial hypothesis. These results have implications for policymakers and political analysts in understanding gender-based electoral trends, which could lead to policymakers targeting these demographics with a different approach in order to maximize the number of votes in upcoming elections.

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

For this assignment, AI tools such as Open AI Chat GPT 2025 were used to assist with the structuring our analysis and explanations. AI helped us draft and make us understand about on potential biases introduced by phone and web surveys, as well as formatting our visualizations, synthesize our initial draft before rewriting, and correct grammar.

However, all analytical processes, statistical modeling, interpretations, and conclusions were entirely made by us. To ensure that the final work reflected our understanding of the course material, we carefully reviewed and edited all AI-assisted sections, cross-referenced them with course readings, class notes, and our knowledge from previous courses. By doing so, we maintained the integrity of our work.

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

9 Appendix

No additional notes are added.