Stratification of political preferences based on gender: Analyzing People's Party Support Using the CES 2019 Phone Survey

STA304 - Winter 2025 - Assignment 2

Shihao Qiu, Kin Fung Ng, Fan Peng, Yunfei Qi

1 Introduction

This report examines whether gender influences support for the People's Party of Canada (PPC) during the 2019 Canadian Federal Election. Previous political science literature has shown that gender can shape voting patterns, with men and women often apparent differences in their support for political parties. By diving into the specific topic of how gender affects support levels for the People's Party of Canada in the 2019 Canadian Federal Election, the report argues that males and females demonstrate distinct voting likelihood patterns.

The significance of this issue rests on understanding how gender affects political preferences, which may have wider implications for party strategies and democratic representation. According to research by Gidengil et al. (2005), Canadian males mostly support right-wing parties while the study by Erickson & O'Neill (2002) reveals that females show an increasing preference for left-wing parties over time. Furthermore, Goodyear-Grant & Croskill (2011) and Bittner & Thomas (2017) share the same view. Men are more likely to support right-wing populist movements, while gender-based perceptions continue to influence political attitudes and choices. Therefore, the PPC's status as a right-wing populist party is essential to learn electoral behaviour by studying gender-based support.

The specific goals of this report are as follows: First, calculate the overall proportion of survey respondents who supported the PPC and construct a 95% confidence interval to assess the precision of this estimate. Second, using logistic regression modelling to evaluate whether gender is a statistically significant predictor of PPC support. Finally, interpreting these results in light of the existing literature on gender and political behaviour.

This report uses data from the 2019 Canadian Election Study(CES) phone survey, with approximately 4,021 Canadians respondents after the federal election. The CES dataset includes demographic and political preference information that helps to investigate how gender affects

political opinions. By addressing this research question, the report aims to provide useful insights to political parties, academics and policymakers interested in how gender affects electoral preferences and how campaigns can better appeal to Canada's diverse voter groups.

2 Data

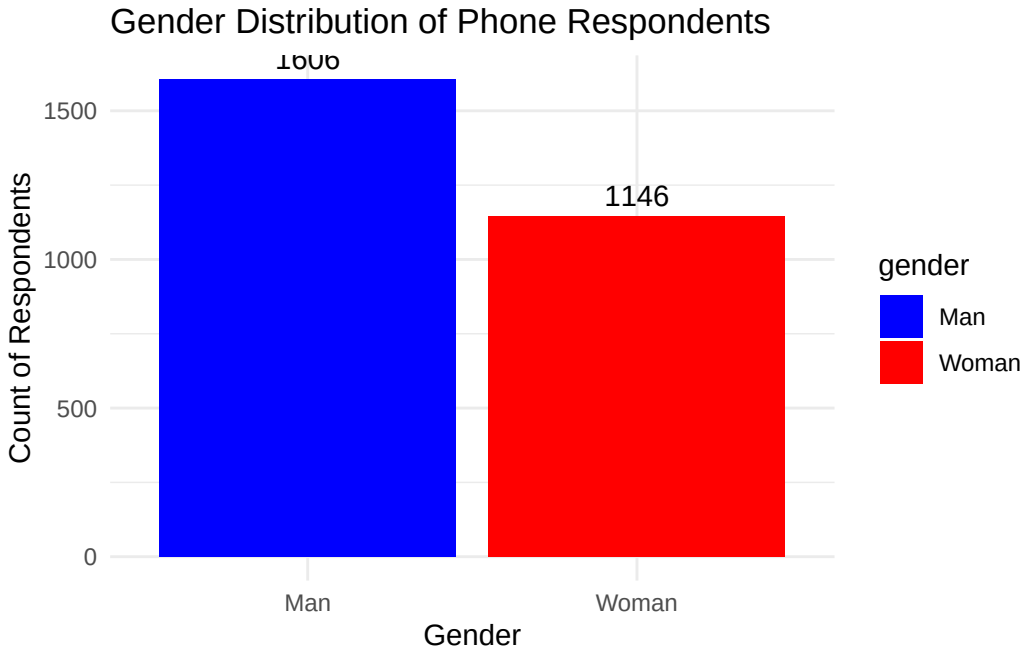
The dataset comes from the 2019 Canadian Election Study (CES) Phone Survey. It collected responses from approximately 4,021 Canadian voters following the federal election. It provides valuable information on voting behaviour, political preferences, and demographics such as education and province of residence. For this report, it focuses on key variables including Vote Choice (`q11`), Gender (`q3`), and Age (`age`). Vote Choice identifies which political party the respondent supported, Gender classifies respondents as Male, Female, or Other, and Age records the respondent's age. We use Vote Choice to identify PPC supporters and Gender as the stratification variable. Age is included as a control variable in later regression analysis.

Table 1: Table 1: Gender Distribution Before Cleaning

gender	count
Man	2272
Woman	1748
Other	1

Gender serves as the stratification variable, as it allows us to separate the data into male and female groups to compare support for the People's Party of Canada (PPC). The `q3` variable categorizes gender responses as 1 = Male, 2 = Female, and 3 = Other. For this report, we exclude the "Other" group due to its very small sample size, which could lead to unreliable estimates if analyzed separately. Stratifying by gender enables us to explore whether there is a significant difference in PPC support between men and women.

To ensure data quality, we conducted a thorough cleaning process. We recorded the `q3` variable into "Man," "Woman," and "Other" for clarity, removing any non-valid gender responses. Additionally, we cleaned the `q11` variable to retain valid responses (1 to 8) and excluded non-responses such as "Refused to answer," "Skipped," or "Don't know." We also created a binary variable to indicate PPC support (1 = voted for PPC, 0 = did not vote for PPC). This data preparation ensures accuracy in our subsequent statistical analysis.



A bar plot was generated to display the gender distribution of respondents within the CES phone survey dataset. The visualization shows that respondents are split between "Man" and "Woman," with a higher number of male respondents. It shows more male respondents are participated in the phone Survey. The "Other" category has no representation after cleaning, confirming its exclusion from the stratified analysis. This distribution confirms that the stratification will be based on male and female groups only.

3 Methods

To estimate the proportion of votes for the People's Party of Canada in the stratified sample, we need to calculate the 95% confidence interval (CI) for proportions by using the following formula:

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

(NCSS, n.d.)
where:
-$\hat{p}$ is the estimated proportion of respondents supporting People's Party of Canada.
-$z_{\alpha/2}$ is the critical value corresponding to the desired confidence level.
-$W_h = \frac{N_h}{N}$ represents the proportion of the population in stratum h (based on gender).
-n_h is the sample size for gender stratum h.

-N_h is the population size for gender stratum h.
-s_h^2 is the sample variance within gender stratum h.

This formula accounts for stratification by weighting each gender stratum based on its proportion in the total population and adjusting for finite population correction $\left(1 - \frac{n_h}{N_h}\right)$. In this formula, Total Population (N) refers to the overall population of eligible voters in Canada who could have been sampled in the 2019 CES phone survey, stratified by gender. This stratified confidence interval calculation improves precision by accounting for gender-based differences in People's Party of Canada support while reducing sampling bias compared to a simple random sample.

To analyze the likelihood of voting for People's Party of Canada based on gender, we apply a logistic regression model:

$$\log\left(\frac{P(Y=1)}{1-P(Y=1)}\right) = \beta_0 + \beta_1 X_{gender} + \beta_2 X_{age}$$

where:
-Y is a binary outcome variable (1 = voted for People's Party of Canada, 0 = did not vote for People's Party of Canada).
-X_1, X_2 are predictor variables (gender, age).
-β_0 is the intercept term, representing the log-odds of voting for People's Party of Canada when all predictors are 0.
-β_1, β_2 are regression coefficients indicating the effect of each predictor on the log-odds of voting for People's Party of Canada.

The intercept β_0 represents the baseline log-odds of voting for the People's Party of Canada when all predictor variables are zero. The coefficient for gender, β_1, is the effect of gender on the log-odds of supporting the People's Party of Canada. A positive β_1 suggests that being male increases the likelihood of voting for the People's Party of Canada, whereas a negative β_1 indicates that being male decreases the probability of support. To interpret this effect in terms of odds ratios, the exponentiation of β_1, given by e^{β_1}, provides the factor by which the odds of supporting the People's Party of Canada change for males compared to females. $beta_2$ has a similar interpretation to $beta_1$

This logistic regression model is well-suited for analyzing binary outcomes. It allows estimates of probability differences related to gender. In addition, by employing a stratified design, the analysis ensures that gender representation in the survey is consistent with the overall population distribution, thereby reducing bias and improving the generality of the survey results. Due to the stratified survey design adopted in this study, we use weighted logistic regression to control gender survey weights in logistic regression.

4 Results

To estimate the proportion of respondents supporting the People's Party of Canada, we computed a stratified confidence interval. This method ensures that variations in support levels across gender are appropriately weighted.

	Proportion of Outcome Variable	95% Confidence Interval of Outcome Variable
Phone Survey	0.012	(0.009, 0.016)

Table 1: The proportions and 95% confidence intervals of outcome variable of interest calculated for the Canadian Election Study 2019 phone survey data.

*For the population estimate "Man" = 18715511 and "Woman" = 18902984, it cited from Statistics Canada (2019).

The percentage of respondents in the 2019 CES telephone survey who supported the People's Party of Canada was estimated at 0.012(1.2%) with a 95% confidence interval of (0.009,0.016). This confidence interval suggests that the true percentage of PPC supporters in the population is between 0.9% and 1.6% with 95% certainty. The relatively small margin of error suggests a stable estimate, although the low overall proportion confirms that support for PPC was not widespread at the time of data collection. This underlines the PPC's status as a minor party in the 2019 general election. The confidence interval also indicates that even under the most favorable conditions (the upper limit of 1.6%), support for the People's Party of Canada remains low relative to the major parties. However, the lower limit (0.9%) suggests support may even be lower than the point estimate. Small sample sizes can be influenced by a number of factors, such as regional differences supported by People's Party of Canada. To further analyze factors influencing People's Party of Canada support, we fit a logistic regression model where the dependent variable is voting for People's Party of Canada (where 1 = voted for People's Party, 0 = did not vote for People's Party). The independent variables include gender and age.

The estimated logistic regression equation for People's Party of Canada support is:

$$log\Big(\frac{\hat{p}}{1-\hat{p}}\Big) = \hat{\beta}_0 + \hat{\beta}_1 x_{gender} + \hat{\beta}_2 x_{age}$$

$$log\Big(\frac{\hat{p}}{1-\hat{p}}\Big) = -4.257710 + 0.374696 x_{gender} - 0.006888 x_{age}$$

Where $\hat{p}$ represents the estimated probability of PPC support.

The intercept -4.258 reflects the baseline log-odds of supporting the PPC for the reference group (female respondents at age zero). Although age zero is theoretical, this baseline helps contextualize how log-odds change with gender and age.

The gender coefficient (0.375) indicates that male respondents have log-odds of PPC support that are higher by 0.375 units compared to female respondents by holding another variable (age) constant. The corresponding odds ratio is $e^{0.375} \approx 1.4545$. It suggests that males are approximately 45% ($(1.454-1)\times 100\% = 45.4\%$) more likely to support the PPC than females. However, the p-value of 0.215 indicates that this gender effect is not statistically significant.

The age coefficient (-0.007) suggests that each additional year of age reduces the log-odds of PPC support by approximately 0.007 units. The corresponding odds ratio is $e^{-0.0069} \approx 0.9931$. It shows that each additional year of age decreases the odds of supporting the PPC by approximately 0.7% ($(0.9931-1)\times 100\% \approx -0.7\%$) by holding gender constant. However, with a p-value of 0.408, this effect is also not statistically significant.

Overall, the results suggest that while there was a slightly higher tendency for men and younger people to support PPC, these patterns were not statistically conclusive. This means that other factors, such as province, education or socioeconomic status, may play a more decisive role. In addition, the baseline probability of PPC gaining support is very low, highlighting the marginal role of PPC in Canada's 2019 election.

5 Discussion

This report set out to explore whether gender influences support for the People's Party of Canada (PPC) during the 2019 Canadian Federal Election. Drawing from the CES Phone Survey data, we aimed to estimate PPC support levels and evaluate if gender plays a statistically significant role in shaping voter preferences. The results showed that only 1.2% of survey respondents supported the PPC, with a 95% confidence interval ranging from 0.9% to 1.6%. While our logistic regression analysis suggested that men were approximately 45% more likely to support the PPC than women, this relationship was not statistically significant because p = 0.215 which is larger than 0.05. Additionally, age showed a minor negative association with PPC support, though this effect was also not significant. These findings confirm that PPC support in 2019 was low and suggest potential gender differences.

Several limitations affect the robustness of our analysis. A major limitation is the exclusion of respondents identified as "other" in the gender variable, as their representation in the data set is small. This decision was made to reduce variance inflation and ensure more stable estimates, but limited the inclusivity of our findings for non-binary individuals. In addition, the model also only uses gender and age as predictors, ignoring important factors such as education, income, and geographic region. The absence of these controls can lead to missing variable biases that limit the explanatory power of the logistic regression models.

These limitations highlight that current models may only capture part of the conclusion that PPC supports. Future research should address these limitations by expanding the model to include a broader range of demographic and socioeconomic predictors. Combining factors such as education, income, political ideology, and geography can deepen understanding of PPC

support patterns. Efforts should also be made to improve sampling and weighting techniques to ensure better representation of underrepresented groups, such as non-binary respondents.

6 Generative AI Statement

During job completion, we used ChatGPT as a complementary tool for formatting aid and Grammarly for syntax correction. Specifically, in Part 3: Methods, we used ChatGPT to convert logistic regression formulas to LaTeX format. The conversion process allows us to display the equation in a clear format, which enhances readability while making it easier to follow in the report. Grammarly corrects grammatical errors such as incorrect verb tenses. And optimize the sentence structure to make the article read more smoothly. This helps to improve the quality of the reports.

To ensure that the final work represents our personal efforts and understanding, we strictly utilize AI to create the formatting and writing language assistance, while we handle all aspects of data set integration, coding, analysis, and validation ourselves. For example, in Part 3: We wrote logistic regression equations in LaTeX using ChatGPT, but chose the right formula through our own critical thinking and logistic regression lectures. In the 8th week of R code practice, the professor also provided a similar formula writing method. We check the accuracy of the LaTeX format by analyzing the knitting formula display as well as the logistic regression model learned in the lecture. Similarly, Grammarly's role is to refine grammatical structure and sentence structure while preserving the original meaning and logical integrity of our analysis. This ensures that AI does not make any analytically decisions, but is instead used as an adjunctive tool to improve the clarity and efficiency of our work.

7 Ethics Statement

To ensure reproducibility, we maintained transparency and clarity throughout the research process. The dataset used which is the 2019 Canadian Election Study (CES) Phone Survey. It is publicly available and accessible to anyone. We documented each step of the analysis, including data cleaning, variable selection, and the application of statistical techniques such as stratified confidence intervals and logistic regression. Our R code was written in a clear, reproducible manner, with comments provided for each significant analytical step. This allows for independent verification and replication of the analysis by other researchers.

Regarding Research Ethics Board (REB) considerations, the analysis was conducted on secondary data that is fully anonymized and in the public domain. As highlighted in class materials, analyses using non-identifiable and pre-existing datasets typically fall under REB exemption guidelines. The CES dataset does not contain personally identifiable information (PII) such as names or addresses, thereby protecting the privacy and confidentiality of survey

participants. Since the data was collected and anonymized prior to our analysis, no additional ethics approval is required for this project.

By adhering to ethical principles of confidentiality, informed consent (via the original CES study design), and data stewardship discussed in the ethics lecture. We ensure that participant privacy is respected and that the findings can be disseminated publicly without compromising human subjects' rights.

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. Bittner, A., & Thomas, M. (2017). Explicit gender stereotyping in Canadian politics: Examining public beliefs about women's suitability for political office. Canadian Journal of Political Science. https://www.cambridge.org/core/journals/canadian-journal-of-political-science-revue-canadienne-de-science-politique/article/explicit-gender-stereotyping-in-canadian-politics/5177A1EDF6BCF17D92EF414E0F219182

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

4. Erickson, L., & O'Neill, B. (2002). The gender gap and the changing woman voter in Canada. International Political Science Review. https://journals.sagepub.com/doi/pdf/10.1177/0192512102023004003

5. Gidengil, E., Hennigar, M., Blais, A., & Nevitte, N. (2005). Explaining the gender gap in support for the new right: The case of Canada. Comparative Political Studies. https://journals.sagepub.com/doi/abs/10.1177/0010414005279320

6. Goodyear-Grant, E., & Croskill, J. (2011). Gender affinity effects in vote choice in Westminster systems: Assessing 'flexible' voters in Canada. Politics & Gender. https://www.cambridge.org/core/journals/politics-and-gender/article/abs/gender-affinity-effect-in-vote-choice-in-westminster-systems-assessing-flexible-voters-in-canada/CB7A26B77B693EE12B5ACB90B4A33E6

7. Government of Canada. (2024, September 25). Canada's population estimates: Age and gender. The Daily. https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1710000501&pickMembers%5B0%5D=1.1&pickMembers%5B1%5D=2.3&cubeTimeFrame.startYear=2019&cubeTimeFrame.endYear=2019&referencePeriods=20190101%2C20190101

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

9. NCSS (n.d.) Confidence intervals for one mean in a stratified design. https://www.ncss.com/wp-content/themes/ncss/pdf/Procedures/PASS/Confidence_Intervals_for_One_Mean_in_a_Stratified_Design.pdf

10. Thomas, M., & Bodet, M. A. (2013). Do women get fewer votes? No. Canadian Journal of Political Science. https://www.cambridge.org/core/journals/canadian-journal-of-political-science-revue-canadienne-de-science-politique/article/abs/do-women-get-fewer-votes-no/BFF54690D696CB502A2340401FA660F8