Analyzing NDP Vote Share Using Gender-Stratified Random Sampling of Web Survey Data from the 2019 Canadian Federal Election Study

STA304 - Winter 2025 - Assignment 2

Group Number 73: Usman Khan, Liban Timir, Aaron White

1 Introduction

Elections are a fundamental aspect of democracies, shaping the leadership that makes critical decisions in society. Understanding voter preferences through statistical analysis of electoral trends is essential for political parties to refine their campaign strategies and effectively engage with key demographics. The 2019 Canadian Election Study is a large-scale survey conducted both before and after the election to examine the political opinions and demographics of Canadian voters (Stephenson et al., 2020). Our study focuses specifically on the pre-election web survey which collected data on key factors such as gender, age, province, voting likelihood, and party preference.

This study aims to answer the following research question: How do gender and age influence the likelihood of voting for the NDP for the web survey data? By addressing this question, this analysis will provide insights into some factors shaping NDP voter support. Understanding these dynamics could help the NDP refine its future campaign and voter outreach strategies. Additionally, adjusting for potential gender representation imbalances helps ensure that NDP vote share estimates are not skewed by differences in gender composition across the sample and target population of eligible voters in the 2019 Canadian federal election.

This study highlights the importance of accounting for methodological biases when interpreting survey data. Political polling often assumes representative samples, but differences in survey demographics may systematically influence results. The next section will provide a more in-depth overview of the data.

2 Data

The data analyzed in this report comes from the web survey of the 2019 Canadian Election Study obtained through the cesR package in R. It has 37,822 responses and contains seven variables: the voter's gender, age, province, education level, interest in the election, voting likelihood, and preferred party.

For this analysis, *gender* is treated as the stratification variable. It is structured such that 1 is a Man, 2 is a Woman, and 3 is Other. There are no NA values across the dataset for the Gender variable. A new binary variable *vote_ndp* was created to represent whether a respondent was most likely to vote for the NDP. There were 6,258 NA responses in the survey to the question of what party they'd be likely for. However, these were not removed and treated as if they would not vote for the NDP. Additionally, an *age* variable was created to be one-hundred subtracted by the *cps19_yob* variable.

The barplot below displays the number of web survey respondents by gender. There were over 20,000 women, approximately 5,000 fewer men, and a small number of respondents who identified as another gender.

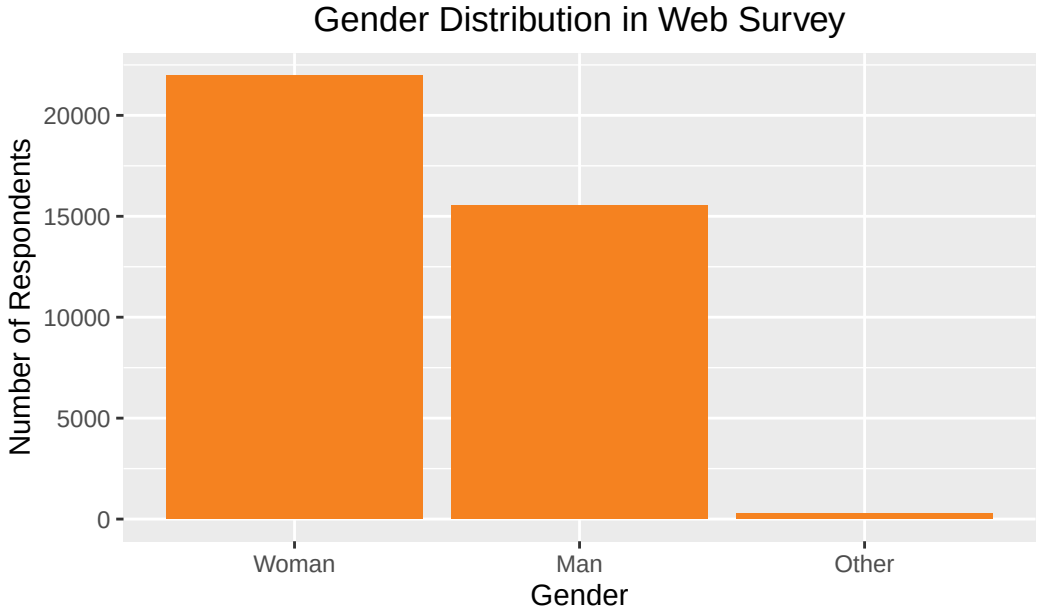

Figure 1. Distribution of gender among respondents in the CES 2019 Web Survey.

3 Methods

In this analysis, we treat the proportion of participants likely to vote for the NDP as the mean of a binary variable (1 if the respondent intends to vote for the NDP, 0 otherwise). We used a

stratified random sampling approach based on gender, dividing respondents into three strata: Man, Woman, and Other. The combined estimator for the overall proportion $\hat{p}_{st}$ is:

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

where H is the total number of strata (3), $\hat{p}$ is the sample proportion in the stratum h, $W_h = \frac{N_h}{N}$ is the proportion of the population in the stratum h, N_h is the population size in the stratum h, and N is the total population across all strata. The total population N_h in stratum h is estimated using national gender population counts from the 2021 Canadian Census (Statistics Canada, 2022).

To assess uncertainty around $\hat{p}_{st}$, we used the stratified random sample confidence interval formula:

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

where $\hat{p_{st}}$ represents the proportion estimate, $z_{a/2}$ represents the critical value for the confidence level, n_h is the sample size in the stratum h, s_h^2 represents the estimated variance of NDP vote preference in the stratum h and $\left(1 - \frac{n_h}{N_h}\right)$ represents the finite population correction factor.

This formula is used to assess the uncertainty around the sample estimate by providing a range that should contain the true population proportion. This interval reflects both the sample variability and the selected confidence level.

To further investigate the factors influencing the likelihood of voting for the NDP, we will use a logistic regression model. In this model, "Male" is treated as the reference category, and we introduce two dummy variables: one for "Female" and one for "Other". The logistic regression model is specified as:

$$\log\left(\frac{P(\text{vote\_ndp} = 1)}{1 - P(\text{vote\_ndp} = 1)}\right) = \beta_0 + \beta_1 \, I(\text{Female}) + \beta_2 \, I(\text{Other}) + \beta_3 \, \text{Age}$$

In this equation, β_0 is the intercept, which represents the log-odds of voting for NDP as a male. β_1 represents the difference in log-odds of voting for NDP between females and males. β_2 represents the difference in log-odds of voting for NDP between other genders and males. β_3 represents the effect of a one-year increase in age on the log-odds of voting NDP. Putting it together, the model examines how gender (with males as the baseline) and age are associated with the likelihood of voting for the NDP.

Together, these two formulas establish a foundation for understanding how gender and age affected the likelihood of voting for the NDP in the 2019 federal election.

4 Results

For the web survey, the estimated proportion of respondents who would be likely to vote for the NDP is approximately 0.1099 with a 95% confidence interval of (0.1068, 0.1131). These results are summarized in Table 1.

Table 1. The proportions and 95% confidence intervals of proportion of votes that the NDP is expected to receive calculated for the Canadian Election Study 2019 web survey data.

	Proportion of Votes that the NDP is Expected to Receive	**95% Confidence Interval of Proportion of Votes that the NDP is Expected to Receive**
Web Survey	0.1099	(0.1068, 0.1131)

The narrow confidence interval (approximately 0.6% in range) indicates that our estimate is highly precise. Given that there are seven political parties to choose from, in an equal distribution of party support there would be a 14.3% expected vote share. Thus, 10.99% expected votes received is below-average support for the NDP.

In Table 2, we present the logistic regression estimates for the web survey from the 2019 Canadian Election Study.

Table 2. Logistic Regression Model Predicting NDP Voting Likelihood for the 2019 Canadian Election Study Web Survey Data

Term	Estimate	Odds Ratio	Standard Error	P-Value
Intercept (Male Gender)	-1.1795	0.3074	0.0560	<0.0001
Other Gender	1.0969	2.9949	0.1350	<0.0001
Female Gender	0.3525	1.4227	0.0351	<0.0001
Age	-0.0239	0.9764	0.0011	<0.0001

The results indicate statistically significant differences in NDP voting likelihood across gender and age. Individuals identifying as not a male or female have 2.99 times higher odds to vote for the NDP compared to males, while female respondents have 42% higher odds compared to males. Additionally, each additional year of age decreases the odds of voting for the NDP by 2.36%. Overall, the findings suggest that older male voters are the least likely demographic to support the NDP.

5 Discussion

When looking at the web survey data for how gender and age influence the likelihood of voting for the NDP, the results were statisticailly significant. We found that people who identified as female or other were more likely to vote for NDP compared to males. We also discovered that younger citizens have incerased odds to vote NDP.

When looking at the web survey data for how gender and age influence the likelihood of voting for the NDP, the results above give us key insights into the real world and allows us to connect other findings to our study. It has been shown that in recent elections that out of all the Canadian parties, women are most likely to be NDP candidates (Medeiros, 2018). In fact during the 2019 federal election the NDP had 49% of candidates who were women, the most out of any party during that election (CBC/Radio Canada, 2019). This might be the reason why we see higher likeliness to vote for NDP if you are a woman, since they are best represented in that party and female citizens may feel that they are better represented (Wicks, 2007). It has been shown that women tend to have an affinity to other female candidates and tend to vote for them more than males (Anderson, 2010). Since the NDP has the most female candidates out of any party, and women are more likely to vote for other women, our results based on the survey may reflect real world patterns seen in elections.

As for the decrease in likelihood of voting NDP as age increases, this could be due to political ideologies differing between generations. Younger generations tend to have more liberal beliefs, like more government aid and higher levels of ethnic and racial diversity (Parker, 2019). This aligns with the policy of the NDP, which outline their initiatives, many of which younger generations tend to gravitate towards (NDP, 2021). This could explain why our study on the survey data shows that as age increases likeliness of voting NDP decreases, since younger generations tend to align more with their values compared to older generations.

In order to improve future research, a few limitations should be addressed. A key limitation is the values that were used for N_h, which were based off the 2021 Census population rather than the actual pool of 2019 eligible voters. This discrepancy may introduce inaccuracies in the stratum weighting process if gender distributions differed between the 2021 Census population and 2019 eligible voters. Furthermore, for the purposes of this assignment, this analysis assumes that the data was collected using a gender-stratified random sampling method. However, the actual sampling methodology that the surveyors used may have differed. Future research into the exact sampling approach used would improve the precision of results.

6 Generative AI Statement

Since our group did not use AI tools, I will explain our workflow process. We started by deciding on a research topic that could be analyzed using the election data. We then analyzed the rubric, and pasted the requirements for each section under their respective titles to ensure

that we cover everything asked of us. We began by writing all the code and calculated all our results, so that we could get a glimpse of what the data entailed regarding our research question. Once we had an idea of how gender and age affected likeliness to vote NDP, we combined our research with our results to make correlations and state our findings within each section.

Our group spent a lot of time planning and studying the requirements needed for this assignment. This meant that all of us had to grasp the course material, so that we could look over each other's work, critique and fine tune our project. We spent a lot of time researching resources outside class material, so that we could make strong arguments backed by academic sources. All of our work has been cited and this assignment strongly reflects our understanding of the course material.

7 Ethics Statement

We ensured reproducibility by following the article Ten Simple Rules for Reproducible Computational Research (Sandve, 2013). Our data manipulation is done by code, we stated every formula we used and we set up public access to our project which includes the code and any updates we add in the future. This ensures reproducibility, since all the data, formulas and code can be accessed here: https://github.com/usman-khan888/STA304-A2.

The data we used does involve human information, but REB approval is not required. In the Elections Canada Privacy statement they mention that they have permission to collect personal information under legislation specifically the Canada Elections Act (Elections Canada, 2024). This would fall under Exception 1, which states that research which relies on data that is public due to legislation does not require REB approval. Also since our research is non intrusive and does not violate the privacy of the participants, there is minimal concern ethically.

8 Bibliography

Anderson, C. D., & Stephenson, L. B. (2010). *Voting behaviour in Canada*. UBC Press.

Canadian Election Study. (2019). 2019 Canadian Election Study. https://www.ces-eec.ca/2019-canadian-election-study/

CBC/Radio Canada. (2019, December 29). *2019 saw a record number of women elected - but gender equity in the Commons is still far off | CBC news*. CBCnews. https://www.cbc.ca/news/politics/women-mps-house-of-commons-2019-election-1.5404800

Elections Canada. (2024, October 17). *Elections Canada and privacy*. https://www.elections.ca/content.aspx?section=abo&dir=pri&document=index&lang=e

Medeiros, M., Forest, B., & Erl, C. (2018). Where women stand: Parliamentary candidate selection in Canada. *Politics, Groups, and Identities*, *7*(2), 389–400. https://doi.org/10.1080/21565503.2018.1557056

NDP. (2021). *Policy of the New Democratic Party of Canada: Effective April 2021*. New Democratic Party of Canada.

Parker, K., Graf, N., & Igielnik, R. (2019, January 17). *Generation Z looks a lot like millennials on key social and political issues*. Pew Research Center. https://www.pewresearch.org/social-trends/2019/01/17/generation-z-looks-a-lot-like-millennials-on-key-social-and-political-issues/

Sandve, G. K., Nekrutenko, A., Taylor, J., & Hovig, E. (2013, October 24). *Ten simple rules for reproducible computational research*. PLOS Computational Biology. https://journals.plos.org/ploscompbiol/article?id=10.1371%2Fjournal.pcbi.1003285

Statistics Canada. (2022). *2021 Census of Population: Canada, provinces and territories – Data tables* (Catalogue No. 98-500-X2021014). Government of Canada. https://publications.gc.ca/collections/collection_2022/statcan/98-500-x/98-500-x2021014-eng.pdf

Stephenson, L. B., Harell, A., Rubenson, D., & Loewen, P. J. (2020). *2019 Canadian Election Study (CES) - Online Survey* (Version 3) [Dataset]. Harvard Dataverse. https://doi.org/10.7910/DVN/DUS88V

Wicks, A., Wicks, Ann, & Lang-Dion, Raylene. (2007). *Equal voice: Electing more women in Canada*. Commonwealth Parliamentary Association, Canadian Region.