Estimations of the Conservatives Support Rate in the 2019 Canadian Federal Election and the Impact of Gender and Age on Support Rate

Qiwei Ye

2025-03-13

1 Introduction

The focus of this report is to analyze the voting patterns of the general public and focuses on the estimated votes won by the Conservative Party in the 2019 Canadian Federal Election, using the online (web) survey from the 2019 Canadian Federal Election Study (Stephenson et al., 2020). In the 2024 US election, we saw a clear split in the political party preferences between young men and women (Yerushalmy, 2024), and a report in Canada found that the support rate of the Conservatives tends to be higher among older people (Coletto, 2025). The report examines whether a similar phenomenon was predicted in the 2019 Canadian Federal election by the survey data. The data was collected via stratified random sampling and the stratification variable is gender. The report aims to estimate the proportion of the voters in the general public that will vote for the Conservatives in the election. It also aims to find the statistical relationship between gender and age and the likelihood of voting for the Conservatives by constructing a logistic regression model.

By completing such analysis, the readers of this report can estimate the expected votes the Conservatives will win in the upcoming election and can tell how gender and age affect political preferences, while in this case, the likelihood of voting for the Conservatives among different gender and age groups. Understanding voter preferences is also crucial for policymakers and political analysts, as these parties can conduct further research to

help understand the causes of the possible differences in voting preferences among different gender and age groups based on the result of this report. Political parties can also develop electoral strategies based on the demographic characteristics of their supporters to win more votes.

The data used in this report is derived from the 2019 Canadian Federal Election Study (CES) web survey data. The CES is a survey designed to capture the attitudes, preferences, and behaviors of Canadian voters during federal elections. By using the 2019 CES data and applying statistical methods, the report can find possible statistical relationships between certain demographic features (in this case, gender and age) and the likelihood of voting for a certain political party (in this case, the Conservatives).

The report aims to estimate the proportion of voters in the general public who will vote for the Conservatives, construct a 95% confidence interval for this proportion, and develop a logistic regression model predicting the logged odds or the logged odds ratio of voting for the Conservatives, using gender and age as the predictors. The report will conclude how are gender and age linearly related to the logged odds of voting for the Conservatives, and interpret the results.

2 Data

The data was collected via stratified random sampling and the stratification variable is gender, which means that the web survey divided the targeted population into strata based on gender, and then randomly drawn samples from each stratum, with the number of individuals drawn from each stratum proportional to the size of the stratum. The dataset records each participant's gender, year of birth minus 1919, province of residency, highest education level, level of interest in the upcoming election, likely to vote, and the political party the participant most likely to vote for.
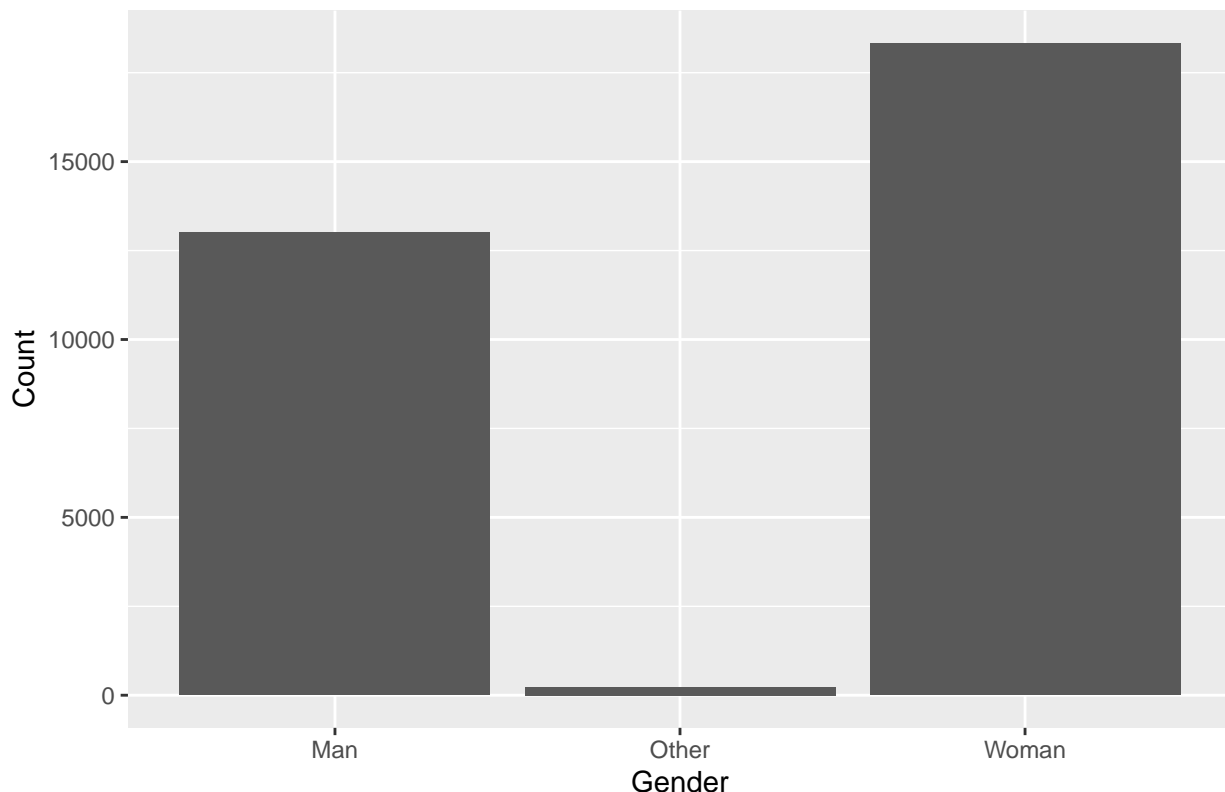
This report is interested in the estimated proportion of votes won by the Conservatives, and how may gender and age affect the likelihood of voting for the Conservatives. Thus, the variables of interest are gender, age, and the party the participant was most likely to vote for.

This report first filtered out any individual who did not provide voting preference data

in the web survey. The original dataset recorded the age for each individual as the year of birth minus 1919 in column "cps19_yob". For simplicity, this report created a new column "age" in the dataset using the mutate() function in R, which equals 100-cps19_yob for each individual, representing the individual's age by the year 2019. The original dataset recorded the gender of each individual in column "cps19_gender" using numerical values, where "1" represents "man", "2" represents "woman" and "3" represents "other". This report created a new column "gender" in the dataset using the mutate() function in R, assigning each individual a verbal description of their gender as "Man", "Woman" or "Other", corresponding to their recorded values in "cps19_gender". The original dataset recorded the political party that the individual is most likely to vote for in column "cps19_votechoice" using numerical values, where each number represents one certain party. This report created a new column "vote_for_conservatives" in the dataset using the mutate() function in R, which is a dummy variable, equals 1 for any individual who responded that they are most likely to vote for the Conservatives, and equals 0 otherwise.

Graph 1 is a bar plot showing the genders represented by the participants of the web survey. The bar plot shows that about 13000 participants of the survey were men, about 18000 were women, and very few identified as "other" compared to men and women.

Bar Plot of Genders Represented by the CES Web Survey Participants

3 Methods

Below is the confidence interval formula used in this report. $\hat{P}$ is the estimated general public proportion of interest, which equals to the sample proportion of interest, in this case, the proportion of the participants of the survey who are planning to vote for the Conservatives in the upcoming election. $z_{\alpha/2}$ is the critical value of a chosen significant level in a normal distribution, as the sample size of the study is big enough for a normal approximation. $\sqrt{\frac{\hat{P}\times(1-\hat{P})}{n}}$ is the standard error of the sample proportion of interest, calculated using the sample size.

$$CI = \hat{P} \pm z_{\alpha/2} \times \sqrt{\frac{\hat{P}\times(1-\hat{P})}{n}}$$

Below is the logistic regression model used in this report. p is the proportion of the general public that is planning to vote for the Conservatives in the upcoming election. β_0 is the interception term, β_1 and β_2 are the coefficients for dummy variable *gender*, which corresponds to $I(gender = Other)$ and $I(gender = Woman)$ respectively. β_3 is the coefficient for variable *age*. The model assumes a linear relationship between the logged odds of voting for the Conservatives, $log(\frac{p}{1-p})$, and the predictors *gender* and *age*.

For the interpretation of the coefficients, the reference level of this logistic regression model is "Man". β_0 represents the logged odds of a 0-year-old man voting for the Conservatives in the upcoming election, β_1 represents the logged ratio between the odds of a non-binary individual voting for the Conservatives in the upcoming election compared to a man of the same age, β_2 represents the logged ratio between the odds of a woman voting for the Conservatives in the upcoming election compared to a man of the same age, and β_3 represents logged ratio between the odds of an individual voting for the Conservatives in the upcoming election compared to a one-year-old younger individual of the same gender.

$$log(\frac{p}{1-p}) = \beta_0 + \beta_1 \times I(gender = Other) + \beta_2 \times I(gender = Woman) + \beta_3 \times age$$

4 Results

Table 1 presents the estimate and the confidence interval of the proportion of voting for the Conservatives. It is estimated that about 27.6% of the general public are planning to vote for the Conservatives in the upcoming election, and the 95% confidence interval is narrow because of the relatively huge sample size, with an upper bound of 28.1% and a lower bound of 27.1%. It is 95% confident that between 27.1% and 28.1% of the general public are planning to vote for the Conservatives in the 2019 Canadian Federal election, using the calculations derived from the data from the 2019 web CES survey.

Estimated Proportion of Voting for the Conservatives	95% Confidence Interval
0.276042	(0.271111, 0.280974)

Table 1: The estimated proportion of voting for the Conservatives and 95% confidence interval calculated for both the Canadian Election Study 2019 web survey data.

Table 2 presents the estimates of the coefficients of the logistic regression model provided in the previous part of this report. For the logistic regression model, the estimated interception is -1.192502, while the remaining estimated coefficients β_1, β_2, β_3 are -0.637675, -0.396522 and 0.009078 respectively.

Based on the estimations from the fitted model, a 0-year-old man has a logged odds of -1.193502 for voting for the Conservatives. When comparing the odds of voting for the Conservatives to the odds of a man of the same age doing so, A non-binary individual has a logged ratio of -0.637075, and a woman has a logged ratio of -0.396522. When comparing the odds of voting for the Conservatives to the odds of an individual who is one year younger and of the same gender doing so, the logged ratio is 0.009078.

In general, non-binary voters and women are less likely to vote for the Conservatives compared to a man of the same age, and older individuals of the same gender on average are more likely to vote for the Conservatives.

	Estimation	P-value
β_0	-1.193502	1.358067e-149
β_1	-0.637675	1.892174e-04

	Estimation	P-value
β_2	-0.396522	1.892174e-04
β_3	0.009078	3.685180e-31

Table 2: The estimated coefficients of the fitted logistic linear regression.

Table 3 presents the estimates on the natural scale by taking the exponential. The results are easier to understand and interpret on the natural scale.

Based on the estimations from the fitted model, a 0-year-old man has an odds of 0.303158 for voting for the Conservatives. When comparing the odds of voting for the Conservatives to the odds of a man of the same age doing so, A non-binary individual has a ratio of 0.528520, and a woman has a ratio of 0.672655. When comparing the odds of voting for the Conservatives to the odds of an individual who is one year younger and of the same gender doing so, the ratio is 1.009119.

	Estimation
e^{β_0}	0.303158
e^{β_1}	0.528520
e^{β_2}	0.672655
e^{β_3}	1.009119

Table 3: The estimated logged coefficients of the fitted logistic linear regression.

It is worth mentioning that the p-values of these estimations are close to zero, which provides strong evidence that the stated relationship between the predictors (gender and age) and the outcome (odds of voting for the Conservatives) exists. For voters of the same age, male on average is the gender that has the highest odds of voting for the Conservatives, while non-binary voters have the lowest odds. Older voters have higher odds of voting for Conservatives compared to younger ones, keeping gender unchanged.

5 Discussion

By using 2019 CES web survey data, through calculation and the use of statistical tools, the report estimated the proportion of votes that the Conservatives are expected to win in the

2019 Canadian Federal election. The report also found the linear relationship of how gender and age affect the logged odds of an individual voting for the Conservatives and transformed the findings to interpret on the natural scale for clarity.

As concluded, the Conservatives are estimated to win between 27.1% and 28.1% votes among all voters. The estimation was done by constructing a 95% confidence interval using the 2019 CES web survey data. By using the variables "age" and "gender" for predicting the odds of an individual voting for the Conservatives through fitting a logistic regression model, this report found strong evidence that among voters of the same age, men have the highest odds of voting for the Conservatives, women have less odds of doing so compared to men, and non-binary voters have the lowest odds. The model also predicts that for voters of the same gender, older individuals have higher odds of voting for the Conservatives.

There are limitations to this report. Firstly, there is potential sampling bias, as the data used for this report is the 2019 CES web survey. The respondents of the survey may not be perfectly representative of the target population, which is all voters in Canada. Not everyone has access to the internet or ever has the chance to take the web survey. Secondly, there is potential social desirability bias when the participants of the survey do not answer the questions honestly. This makes sense when it comes to questions related to political preferences, the participants may select a more socially desirable political party for the survey rather than their true opinion. Thirdly, this report only examined two predictor variables, gender and age, and their impact on the odds of voting for the Conservatives, while the web survey dataset contains many more variables which may have a similar effect on the outcome of interest, those variables are not discussed here.

The result of this report only suggests the correlation between the predictors and the outcome and does not suggest any causal relationship. Future research could focus on examining the causes of the difference in political preference between gender and age groups, which would be very important for political parties as they can better identify and target their supporters. Future research could also include more predictor variables to fit the regression model, or choose a different dataset to test if a different population is reached and how may it change the results concluded by this report.

6 Generative AI Statement

The generative AI (ChatGPT 4o) helped to provide the codes that set the font size and margins of this report, and also helped debug several code chunks. Also, ChatGPT 4o helped find two past online articles about differences in political preferences related to demographic characteristics, introduced in the introduction part of this report. The article ChatGPT 4o provided was read in full by myself. All other work was done by myself independently, without any help from other students or generative AI.

7 Ethics Statement

The analysis of this report was completed in a R Markdown environment. The report cited the data source, described the data cleaning process, and stated all formulas and models used for producing the output. Although the readers cannot directly view the codes that produced these outputs from this document, the attached rmd file contains all the codes that have been used. All calculations were done by codes in R, including the calculation of confidence interval and estimations of the coefficients in the logistic regression model, and strictly used the provided formulas and models. There is no component in this report that contains any random selection or simulation process. The analysis is expected to be reproducible.

The CES 2019 data is human participant data, and the publicly available versions of the dataset are anonymized, all personally identifiable information is removed. The data is publicly available, the work completed in this report does not Link the CES data with any other data that may contain personal information, and does not try to re-identify individuals in the data. The privacy of the participants of the web survey used for the analysis is protected. Based on these circumstances, the work completed in the report does not need Research Ethics Board approval for the report the be made publicly available, according to the criteria of TCPS2 (Government of Canada, Interagency Advisory Panel on Research Ethics, 2023).

8 Bibliography

1. Coletto, D. (2025, January 19). Abacus Data Poll: Conservatives lead by 26 as Liberal leadership election kicks off. Abacus Data. https://abacusdata.ca/canadian-politics-abacus-data-january-2025/

2. Dekking, F. M., Kraaikamp, C., Lopuhaä, H. P., & Meester, L. E. (2005). A modern introduction to probability and statistics. In Springer texts in statistics. https://doi.org/10.1007/1-84628-168-7

3. Government of Canada, Interagency Advisory Panel on Research Ethics. (2023, January 11). Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans – TCPS 2 (2022). https://ethics.gc.ca/eng/policy-politique_tcps2-eptc2_2022.html

4. Grolemund, G. (2014, July 16). Introduction to R markdown. https://rmarkdown.rstudio.com/articles_intro.html

5. OpenAI. (2025). ChatGPT (Version 4o) [Large language model]. https://chat.openai.com

6. Stephenson, L. B., Harell, A., Rubenson, D., & Loewen, P. J. (2020). 2019 Canadian Election Study (CES) - online survey [Dataset]. In Harvard Dataverse. https://doi.org/10.7910/dvn/dus88v

7. Yerushalmy, J. (2024, November 14). What's behind the global political divide between young men and women? The Guardian. https://www.theguardian.com/us-news/2024/nov/14/us-election-donald-trump-voters-gender-race-data