# Are Women More Likely To Vote Liberal?

**STA304 - Winter 2025 - Assignment 2**

GROUP NUMBER: 77

## 1 Introduction

In this section you will briefly describe your report. Explain the importance of the subsequent analysis and prepare the reader for what they will read in the subsequent sections. Provide an overview of the research question. Briefly describe the 2019 Canadian Federal Election Study and its relevance. State the purpose and goals/hypotheses of the report.

Elections are a cornerstone of democracy, and understanding voter behavior is crucial for predicting outcomes and interpreting political trends. Election polls not only reflect public opinion but can also influence voter turnout and vote choice (Dahlgaard, 2017). However, the validity of these polls depends on how data is collected. Phone surveys may oversample certain demographics due to selection bias, while web surveys are prone to coverage errors and self-selection biases (Harrison, 2023). These methodological differences can shape demographic distributions and affect the conclusions drawn from the data.

Knowing the liberals won the election, we would like to gain some insgihts on the demographic breakdown of their votes This study investigates the effect of **gender** on the **intention to vote Liberal**, using data from the 2019 Canadian Election Study (CES). Females tend to have higher intention to vote for liberal parties, given that these parties are generally characterized by support for … . Investigating the relationship between gender and liberal voting intention shows us how socioeconomic factors influence political preferences, particularly in the context of growing ideological polarization.

Overall, by exploring how the distribution in the study's participants' education level and the proportion of people intending to vote for the liberal party can help us assess potential biases introduced by survey methodology. Hence this study contributes to a deeper understanding of voter behavior and the importance of survey design in political research.

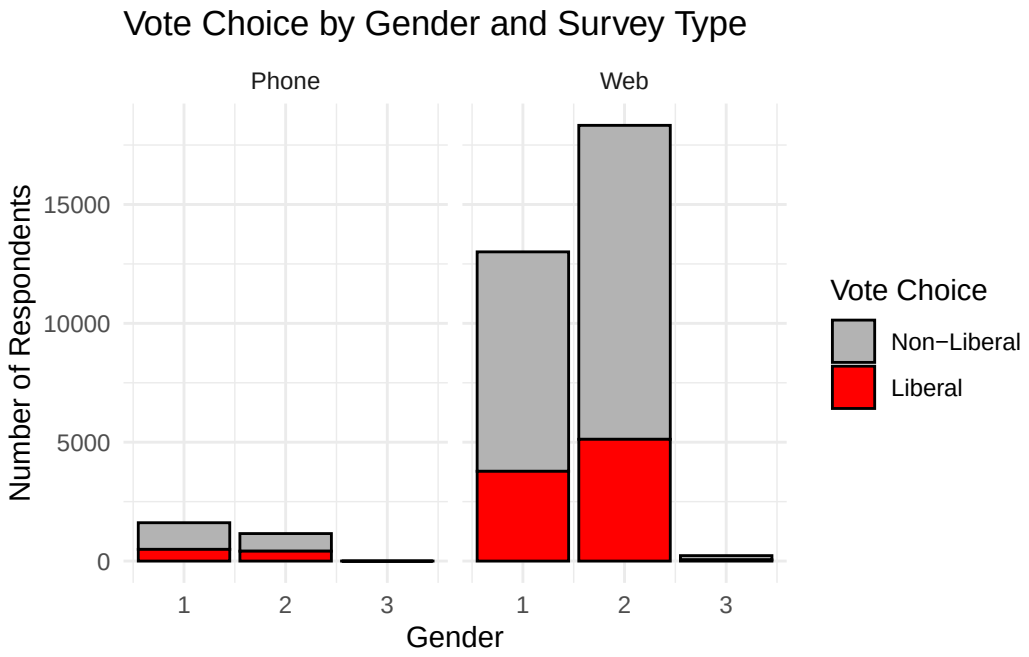
## 2 Data

Briefly introduce the data and key variables of interest. If you do any general data cleaning or data processing you should describe it (in a reproducible manner) here. Identify the stratification variable used. Include at least one plot displaying the distribution of the strata variable. If you do any data cleaning or data processing to the you should describe it (in a reproducible manner) in this section.

The analysis draws from the 2019 Canadian Federal Election Study (CES), which was collected using stratitifed random sampling by gender. In 2019, 50.3% of the Canadian population were women (18.92 million out of 37.6 million). For the CES, data was obtained via phone and web surveys. The participants of each survey are asked identical questions. These questions covered demographic information (e.g., education level and gender), interest in the election, likelihood to vote, and intended vote choice.

In this study, the data cleaning involved two main steps. First, we dichotomized the primary outcome variable, voting intention for the people's party. Second, entries with missing or invalid responses for intended party or gender were excluded. These exclusions were necessary because the analysis looked at differences between this across the two studies and we do not one to inflate one of them by including entries that are missing the other entry.

Be sure to have text describing any plots or tables included.



## 3 Methods

Include the formula for calculating the confidence interval for proportions (do not include specific numbers yet) and provide a description of its components. Present the logistic regression model, specifying the independent variables and expected interpretation of coefficients (parameters, not estimates), and describe the model. In this section you are preparing the reader for how to interpret the numbers displayed in the next section (Results).

Be sure to include your CI formula, and a description of the components of the formula.

$$p \pm Z\alpha/2\sqrt{\frac{p(1-p)}{n}}$$

Be sure to include your logistic regression model, and a description of the components of the model.

$$y = mx + b$$

## 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.

```
calculate_proportion_ci <- function(data, variable, alpha = 0.05) {
  n <- nrow(data)
  proportion <- mean(data[[variable]], na.rm= TRUE)
  se <- sqrt((proportion * (1 - proportion)) / n)
  z <- qnorm(1 - alpha / 2)
  ci_lower <- proportion - z * se
  ci_upper <- proportion + z * se
  list(proportion = proportion, ci_lower = ci_lower, ci_upper = ci_upper)
}

phone_results <- calculate_proportion_ci(phone, "voteliberal")
web_results <- calculate_proportion_ci(web, "voteliberal")

# Display the results
table1 <- data.frame(
  Survey = c("Phone", "Web"),
  Proportion = c(phone_results$proportion, web_results$proportion),
  `95% CI` = c(
```

```
    sprintf("(%.3f, %.3f)", phone_results$ci_lower, phone_results$ci_upper),
    sprintf("(%.3f, %.3f)", web_results$ci_lower, web_results$ci_upper)
  )
)

knitr::kable(
  table1,
  col.names = c("Survey", "Proportion of Right Voters", "95% Confidence Interval"),
  align = "lcc",
  caption = "Proportion of respondents intending to vote for a right-wing party by survey met
) %>%
  kableExtra::kable_styling(
    full_width = FALSE,
    position = "center",
    bootstrap_options = c("striped", "hover")
  ) %>%
  kableExtra::add_header_above(c(" " = 1, "Intention to Vote for People's Party" = 2))%>%
  kableExtra::add_footnote("The table includes 95% confidence intervals for each survey type
```

```
Warning in kableExtra::add_footnote(., "The table includes 95% confidence
intervals for each survey type (Phone and Web)."): Notation is set to 'number'
and other formats are not supported.
```

Table 1: Proportion of respondents intending to vote for a right-wing party by survey method.

| | Intention to Vote for People's Party | |
|---|---|---|
| Survey | Proportion of Right Voters | 95% Confidence Interval |
| Phone | 0.3277978 | (0.310, 0.345) |
| Web | 0.2835192 | (0.279, 0.288) |

Provide a table or formula of the estimated logistic regression model, and include text describing this table/formula and the key takeaways. Interpret the estimates from the logistic regression model. Specifically, commenting on how the predictor variables relate to the outcome variable.

Below is the estimate regression model....

$$\widehat{logit(votepeoples)} = \beta_0 + \beta_1 * gender + \beta_2 * age$$

## 5 Discussion

Summarize key findings. Discuss limitations of the analysis (e.g., potential biases, missing variables, survey errors). Provide recommendations for future research or improvements.

## 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.

Alternatively, if you did not use Generative AI, please include a brief statement outlining your workflow for completing this assignment.

## 7 Ethics Statement

Explain how you ensured that your analysis is reproducible (e.g., documenting code, using proper statistical methods).

Since the CES 2019 data is publicly available, describe whether or not this the work completed in your report needs Research Ethics Board approval for the report the be made publicly available. Be sure to specifically discuss the privacy of human participants in this study.

# 8 Bibliography


1. Grolemund, G. (2014, July 16) *Introduction to R Markdown*. RStudio. https://rmarkdown.rstudio.com/articles_intro.html. (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. Allaire, J.J., et. el. *References: Introduction to R Markdown*. RStudio. https://rmarkdown.rstudio.com/docs/. (Last Accessed: April 4, 1991)

# 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.