Advising the liberal party to target specific group of voters primarily based on voters' education background

Chloe Huang.

December 7, 2020

Introduction

Using the data from the 2019 Online Canadian Election Survey, which investigates 37,822 people eligible to vote, we will advise the **liberal party** to focus on **specific group of voters with a specific education background** by exploring 3 questions:

1. whether **more educated** (voter with at least Bachelor's degree) and **less educated** voters (voter without Bachelor's degree) in Canada have similar average rating score for the liberal party,

2. whether the proportion of **less educated men and women** in Canada who think the government should spend more on education is similar,

3. how **age of less educated women in Alberta** is associated with their rating score for Justin Trudeau, the liberal party leader.

**Note that "voters" refer to people in Canada who are eligible to vote.

Data summary

Before exploring the 3 research questions, we will first clean and organize our data roughly:

- Remove those who don't share about their education background.
- Categorize voters with at least Bachelor's degree as **More educated** and voters without Bachelor's degree as **Less educated**.
- Keep only factors/variables related to research questions (see Table 1 on the next page for detailed variables)

Table 1: Introduction of Variables used

Variable Explanation	**Use in**
Voter's education Background: More / Less educated	Q1,Q2,Q3
Voter's rating score for the liberal party: 0-100	Q1
Voter's gender: Man, woman, or other	Q2
Voter thinks education should be invested more/less/same	Q2
Voter's age: Ranges from 18-99	Q3
Province: Where the voter currently lives in	Q3
Voter's rating score for the liberal party leader: 0-100	Q3

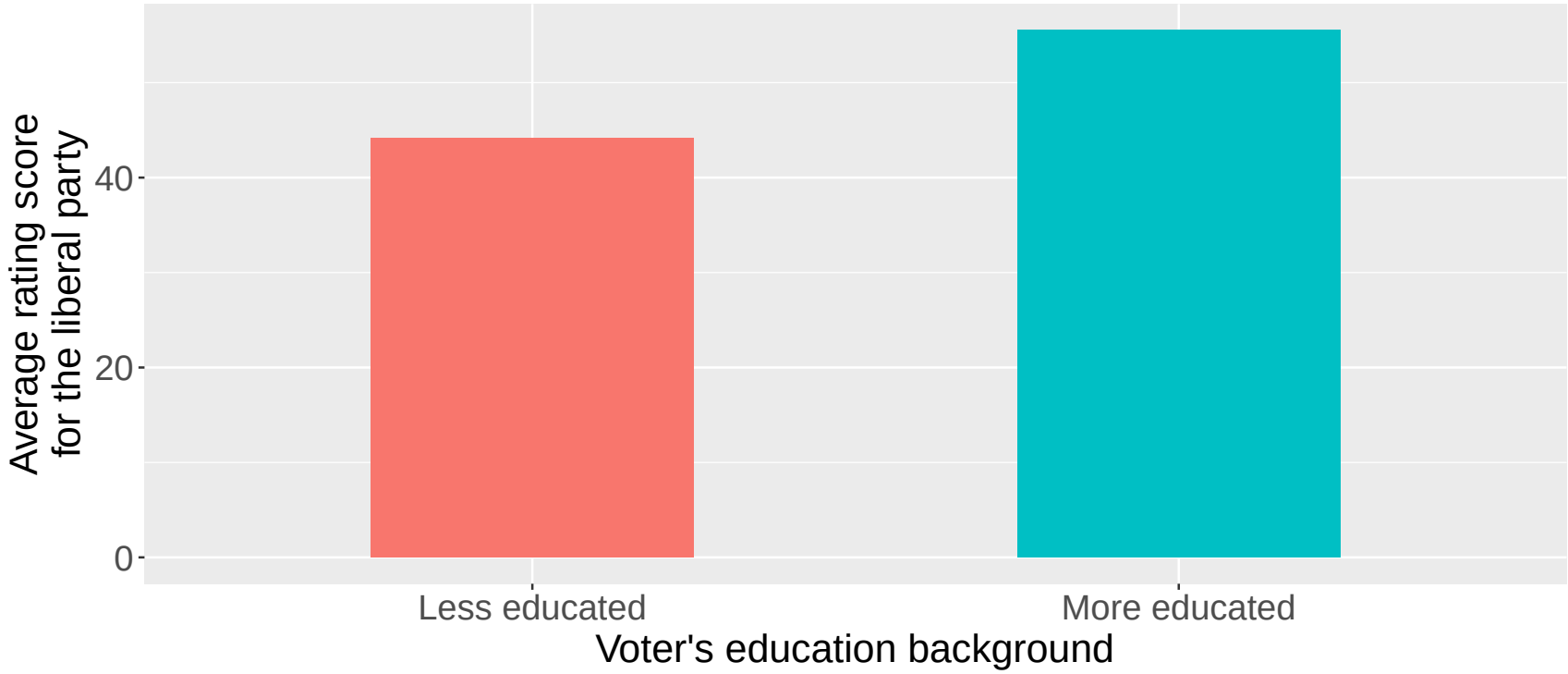
Note that this is only a rough data cleaning. As each question requires different combination of data, we will further organize data before handling each question.

Q1: Do more educated and less educated voters have similar average rating score for the liberal party?

❶ Data cleaning & Visualization

We only pick voters who rate for the liberal party and construct a bar plot that illustrates the average rating score by more/less educated group.

Figure 1: Average rating score (out of 100) for the liberal party by more educated and less educated group



2 Method

Aim: We apply **hypothesis test** to compare whether average rating score between more and less educated groups is similar or not.

1. Two hypotheses for hypothesis test:

- **Null Hypothesis** (H_0): There is no difference in average rating score for the liberal party between more & less educated group.

$$H_0 : \mu_{more.educated} - \mu_{less.educated} = 0$$

- **Alternative Hypothesis** (H_1): There is difference in average rating score for the liberal party between more & less educated group.

$$H_1 : \mu_{more.educated} - \mu_{less.educated} \neq 0$$

  where $\mu_{more.educated}$ and $\mu_{less.educated}$ denote more and less educated group's average liberal party rating score, respectively.

- Our aim is to determine **whether there is enough evidence against the null hypothesis**.

2. Calculate Test Statistic

- The test statistic, which is the difference in average liberal party rating score between two groups based on our data, is 11.37. As figure 1 shows less educated group has lower average rating score, less educate group's average rating score is 11.37 lower than more educated group.

3. After using R to simulate how values would distribute if there is no difference between two groups, we compare test statistic with simulated values to get a p-value (i.e. a value that shows the proportion of simulated values that are at least as extreme as test statistic).

3. **Result & Interpretation**

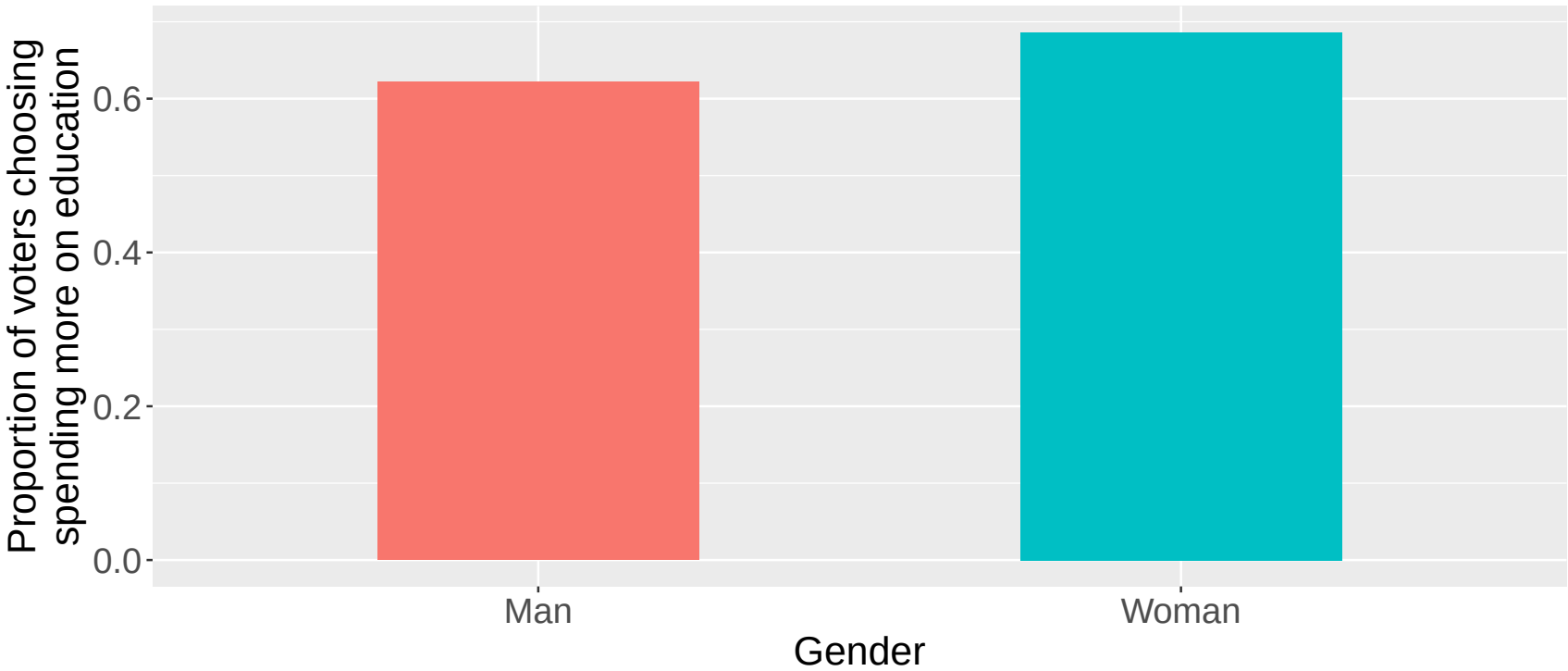
- The p-value is 0, meaning that the test statistic (difference in the sample) has 0% chance to occur if there is no difference between two groups.
- Hence, there is very strong evidence against the null hypothesis that there is no difference in average liberal party rating score between more & less educated group.
- As we know less educated group has lower average rating from figure 1, we have very strong evidence that less educated group has **lower** average liberal party rating score than more educated group.

Q2: Is the proportion of less educated men and women who think the government should spend more on education similar?

❶ Data cleaning & Visualization

Selecting only less educated men and women, we construct a bar chart to illustrate the proportion of less educated voters choose spend more on education in each gender group.

Figure 2: Proportion of less educated voters choosing spending more on education in each gender group



2 Method

We apply **hypothesis test** to compare if proportion of voter who choose spend more on education between less educated men & women are similar.

1. Two hypotheses for hypothesis test:

- **Null Hypothesis** (H_0): There is no difference in proportions of less educated men and women who choose government should spend more on education.

$$H_0 : p_{men} - p_{women} = 0$$

- **Alternative Hypothesis** (H_1): There is difference in proportions of less educated men and women who choose government should spend more on education.

$$H_1 : p_{men} - p_{women} \neq 0$$

  where p_{men} and p_{women} denote the proportion of men and women who choose government should spend more on education, respectively

- Our aim is to determine **whether there is enough evidence against the null hypothesis**.

2. Calculate Test Statistic

- The test statistic, which is the difference in proportions choosing spend more on education between two groups based on our data, is 0.065. As figure 2 shows women has higher proportion, proportion of less educated women who think government should spend more on education is 0.065 higher than proportion of less educated men.

3. After using R to simulate how values would distribute if there is no difference between two groups, we compare test statistic with simulated values to get a p-value.

3. **Result & Interpretation**

- The p-value is 0, meaning that the test statistic (difference in the sample) has 0% chance to occur if there is no difference between two groups.
- Hence, there is very strong evidence against the null hypothesis that there is no difference in proportion of less educated men and women who choose spend more on education.
- As we know women have higher proportion than men from figure 2, we have very strong evidence that proportion of less educated women who think government should spend more on education is **higher** than men.

Q3: Is the age of less educated Alberta women associated with their rating score for Justin Trudeau, the liberal party leader?

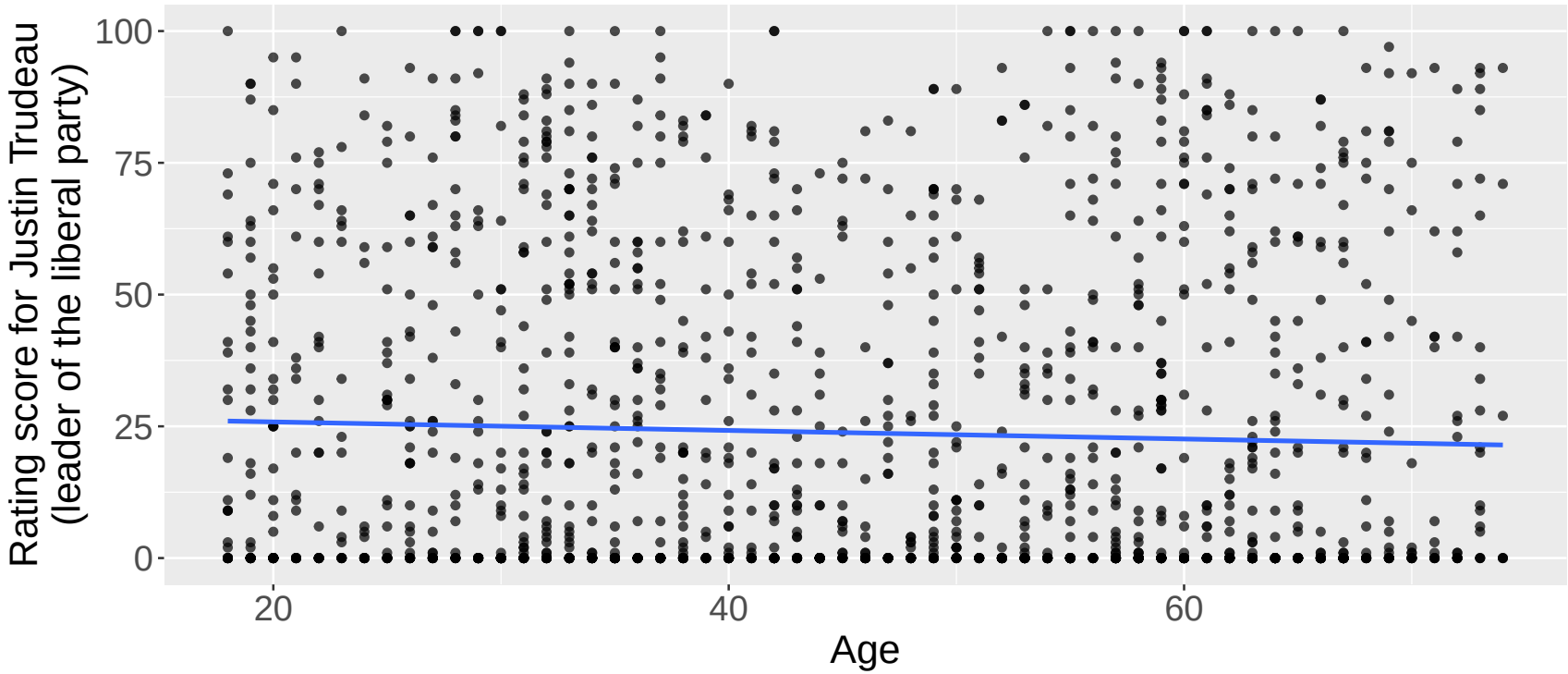
1. **Data cleaning & Visualization**

We only pick less educated women who

- are living in Alberta
- rate for Justin Trudeau (the leader of the liberal party)
- age is below 75 (since too few data above 75 years old; if included, significantly alters the association).

Then we construct a scatter plot (on the next page) to illustrate the association between less educated Alberta woman's age and their rating score for Justin Trudeau.

Figure 3: A scatter plot of less educated Alberta woman's rating score for Justin Trudeau versus their age



2 Method

- We apply **simple linear regression** to figure out the association.
- The regression model is

$$Rating_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

where $Rating_i$ is *ith* Alberta woman's rating for Justin Trudeau. β_0 is the mean rating by 0-year-old woman. β_1 is the change in mean rating score when Alberta woman's age increases by 1. x_i is *ith* Alberta woman. ϵ_i is the difference between *ith* woman's rating score and the mean rating score predicted by the model.

- Our aim is to (1) figure out the **fitted regression equation**, or estimate β_0 and β_1 based on our data, (2) test if β_1 is 0, or whether less educated Alberta woman's **rating score indeed changes as age changes** by using hypothesis test:

  Null Hypothesis: $H_0 : \beta_1 = 0$ (rating doesn't change as age changes)

  Alternative Hypothesis: $H_1 : \beta_1 \neq 0$ (rating changes as age changes)

  We will determine if there is enough evidence against the null hypothesis.

3 Result & Interpretation

- Based on the data we have, we estimate β_0 to be 27.82, β_1 to be -0.10, meaning that mean rating for Justin Trudeau decreases by 0.10 when less educated Alberta woman's age increases by 1.
- Thus, the fitted regression equation based on our data is

$$\hat{Rating}_i = 27.82 - 0.10x_i$$

  where x_i is the age of *ith* less educated Alberta woman.

- R automatically calculates the p-value for β_1, which is 0.047, meaning that we have moderate evidence against the null hypothesis which states the slope parameter β_1 is 0, or less educated Alberta woman's rating for Justin Trudeau is not associated with age.
- Together with figure 3, which shows the negative association, we conclude there is moderate evidence that less educated Alberta woman's rating score for Justin Trudeau **decreases as age increases**.

Conclusion

Table 2: Results and their strength

Question	Result	Strength
1	Less educated voters have lower average rating score for the liberal party than more educated voters	Very Strong
2	Among less educated voters, higher proportion of women think the government should spend more on education	Very Strong
3	As less educated Alberta woman's age increases, rating score for Justin Trudeau decreases	Moderate

The three questions go from general to specific. We recommend Liberal Party to

1. Generally...
   - **target less educated voters** (voters without Bachelor's degree) more than more educated voters
2. Specifically...

Limitations

1. From an overall perspective...

We primarily focus on less educated group, who are less in favor of the liberal party. Yet, it would be beneficial to further explore how **more educated people of different identities** (e.g. gender/religion/age...) think about the liberal party, and figure out how to maintain their relatively high approval rate.

2. Regarding specific research method...

The linear regression method we used in question 3 assumes there is linear association between less educated Alberta woman's rating for Justin Trudeau and their age. However, figure 3 shows this linear relationship is loose. Hence, **linear regression might simplify the relationship as linear and may not reflect the actual association**.