STA130 Final Project

Investigating rating of the Liberal Party from three social perspectives

Project Group 45: Harriet Zhu (1006678150), Yichen Jiang (1006851355), Runshi Yang (1006734223), Yu Zhang (1006764592)

December 7, 2020

Introduction

In this project, we will be giving advice to Canada's **Liberal Party**, which is the federal political party that currently rules the country. In order to do this, we would analyzed variables from the data *the 2019 Online Canadian Election Survey*. Our suggestions will based on the ratings of the Liberal party from raters categorized by three different perspectives: education, job and sexuality.

i) **Education**: how would individuals who have bachelor's degree or above rate Liberal party.

ii) **Job**: whether there is a difference in the rating of the party between the working individuals and people who have already been retired.

iii) **Sexuality**: whether there is an association between sexuality of raters and rating of Liberal party.

Objectives

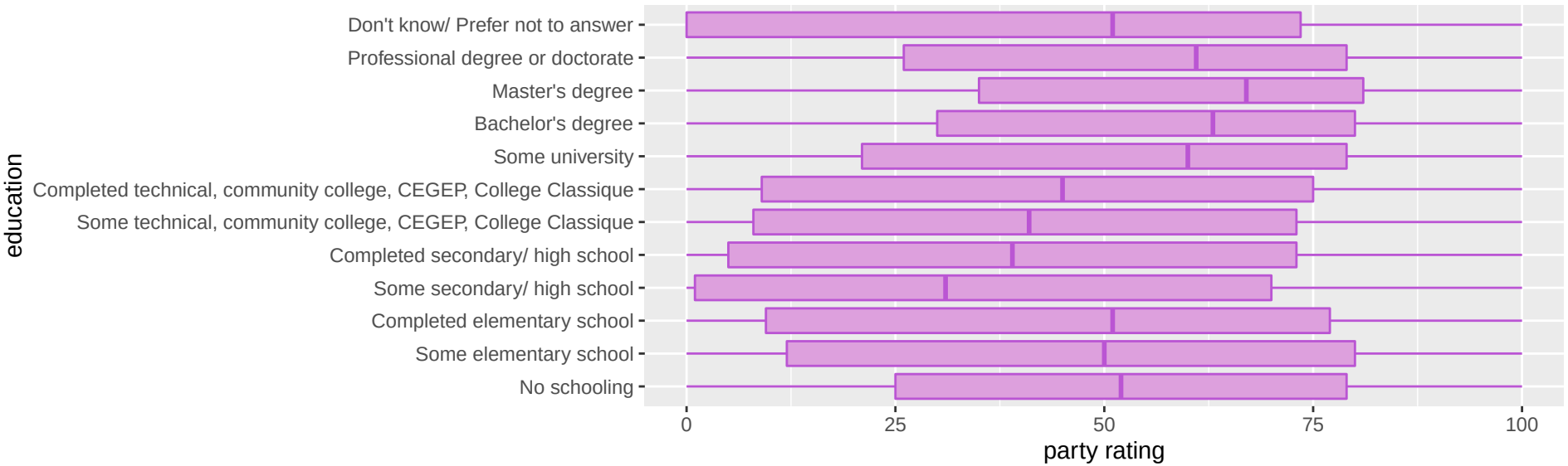
Specifically, these are the **three** questions that we are interested in:

1. What is **the range of plausible values** for the average rating of Liberal party from individuals with a bachelor's degree or above?
2. Is the average rating of Liberal Party **similar** between raters aged 18 to 65 and raters aged 65 to 99?
3. Is there an **association** between sexuality of raters and rating of Liberal party?

Question 1: Data Summary

For Q1 (What is the range of plausible values for the average rating of Liberal party from individuals with a bachelor's degree or above?), we cleaned the data by removing all the NAs in *education* and *party_rating_23* variables. Also, by filtering education, we found that there are 13069 out of 35176 Canadian voters who had bachelor's degree or above.

Fig1. Rating of Liberal party for different levels of education



We are interested in the rating of individuals with bachelor's degree or above because their median ratings are relatively higher from *fig1*.

Question 1: Statistical Methods

Our sample is the individuals who rated for Liberal party who had bachelor's degree or above.

Mean rating of the Liberal Party from individuals with a bachelor's degree or above

```
## # A tibble: 1 x 1
##   `mean(party_rating_23)`
##                     <dbl>
## 1                    55.5
```

We first calculate the sample statistic of 55.507 by getting the mean rating of 13069 individuals who had bachelor's degree or above. Then, we use a **bootstrap method** to investigate the problem. This means that we take out many bootstrap samples of size of the observations (n = 13069), with replacement, from the original sample.
We then simulate 1000 bootstrap samples and calculate the average party rating (Liberal Party) of individuals with a bachelor's degree or above for every bootstrap sample.

Question 1: Results

We calculate a 95% confidence for our sampling distribution of the mean rating for the Liberal party from individuals who had bachelor's degree or above.

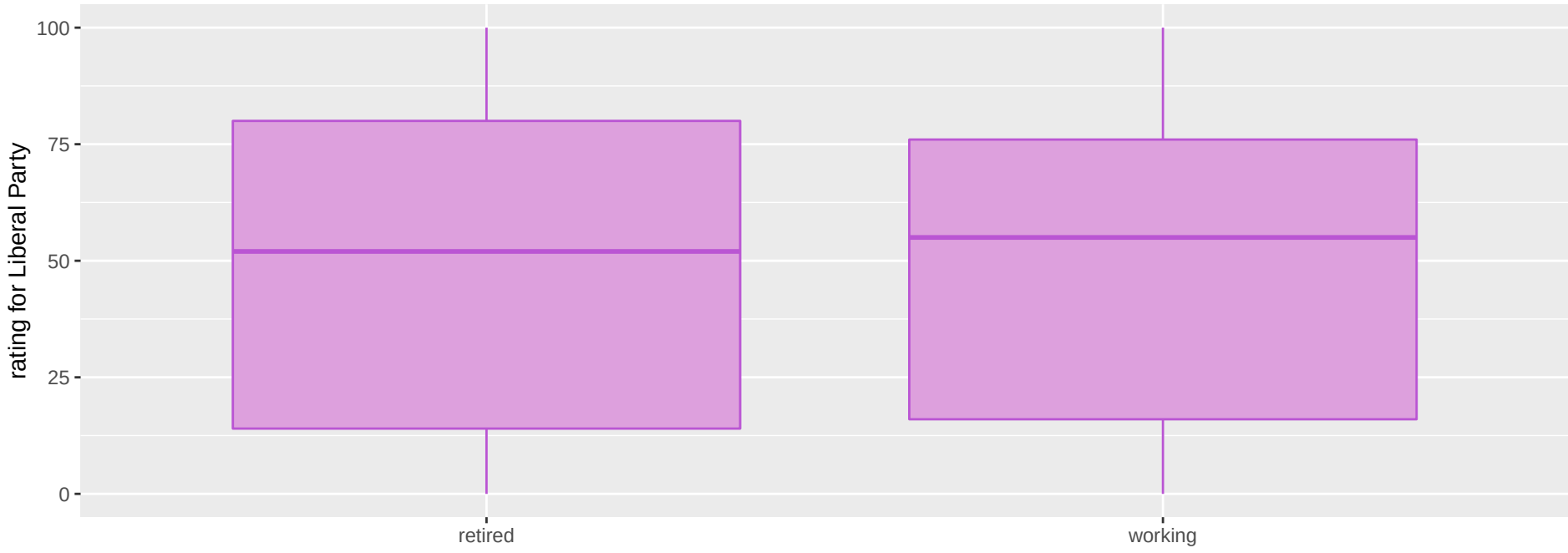
```
##      2.5%     97.5%
## 55.01329 56.03571
```

The result shows that we're 95% confident that the mean average party rating (Liberal Party) of raters with a bachelor's degree or above is between 55.01329 and 56.03571, which means raters with bachelor's degree or above tend to give a higher rating than others, compared to the overall average party rating of 48.3625.

Question 2: Data Summary

We cleaned the data by removing all the NAs in *age* and *party_rating_23* variables for Q2 (Is the average rating of Liberal Party similar between raters aged 18 to 65 and raters aged 65 to 99?). Since the minimum retirement age of Canada is 65, we created a variable called *age_group* and set the value to 'working' for the observations between 18 and 60 and 'retired' for those above 65. Then, we plotted a set of boxplots (*fig2*) to show the distributions of ratings of the two groups.

Fig2. Rating of Liberal party for 'working' and 'retired' groups



Question 2: Statistical Methods

We use a **two-sample hypothesis test** to compare the average rating of the party between the two groups. The null and alternative hypotheses are listed below:

- $H_0 : \mu_{working} - \mu_{retired} = 0$: there is no difference in the average rating of Liberal party between people who are working and retired individuals.
- $H_1 : \mu_{working} - \mu_{retired} \neq 0$: the average rating of Liberal party is different between people who are working and retired individuals.

To get the test statistic, we calculate the difference between the average rating in two groups, and then simulate under the null hypothesis, that is, shuffle the values to get the distribution of the differences in the average rating of two groups.

Question 2: Results

Finally, we evaluate the evidence against the null hypothesis and make a conclusion based on the p-value calculated.

P-value calculated

```
## [1] 0.983
```

We have a **p-value** of 0.983 that 98.3% of our simulations under the null hypothesis were same or more extreme than our test statistic.

This means we have no evidence against the null hypothesis that there is no difference between the mean rating between the working and retired people, which indicates that **Liberal party's rating did not differ between working people and retired people**.

Question 3: Data Summary

For Q3 (Is there an association between sexuality of raters and rating of Liberal party?), we also removed all the NAs in *party_rating_23* variable and "Prefer not to say""Don't know" in *sexuality* variable. We then created a variable called *sexuality_group*. We set observations with Heterosexual sexuality into "heterosexual" category and observations with other sexualities into "others" category because heterosexual is still a major sexuality. Then, we picked out the variables we needed.

A summary table of no. of observations in different sexuality groups and their median ratings of the Liberal Party.

```
## # A tibble: 2 x 3
##   sexuality_group     n median_rate
##   <chr>           <int>       <dbl>
## 1 heterosexual    30387          52
## 2 others           3418          63
```

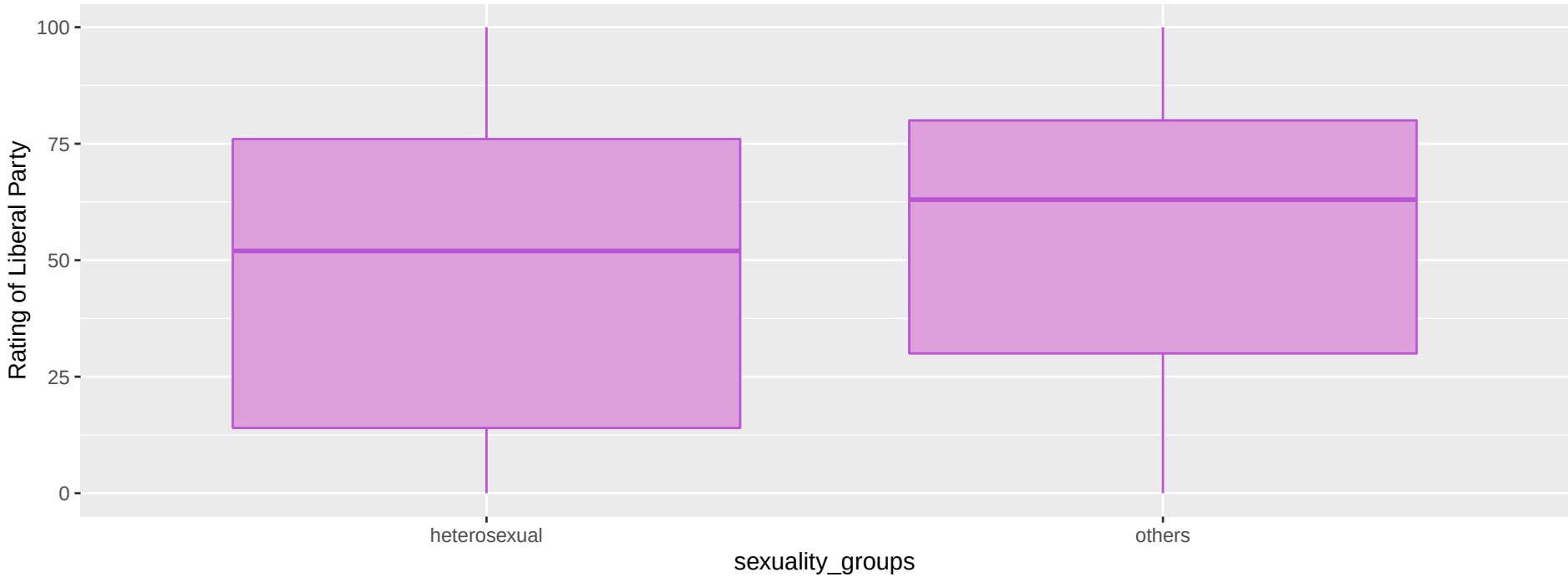
Question 3: Statistical Methods

We use a **simple linear regression model** to predict the association between sexuality groups and average rating of Liberal Party. Since the sexuality group is a categorical variable, this fitted linear regression model has a categorical predictor.

$$LPsupportrate_i = \beta_0 + \beta_1 group_i + \epsilon_i$$

- $LPsupportrate_i$ is the **response variable** which means the rating of Liberal Party of ith selected individual.
- $group_i$ is the **predictor variable** which stands for the sexuality groups of voters. The baseline value is group 'heterosexual' = 0 and group 'other' = 1
- β_0 is the **y-intercept** of the linear regression model and the **estimated mean rating** for Liberal party from the heterosexual group.
- $\beta_1$is the **gradient** of the modal and **average difference** in the response variable y between two groups.

Fig3. Rating of Liberal party between different sexuality groups



After seeing the differences in *fig3*, we then carry out a **hypothesis test** for our fitted regression model:

- $H_0 : \beta_1 = 0$: there is no difference in rating of Liberal Party between heterosexual group and group of other sexualities.
- $H_1 : \beta_1 \neq 0$: there is a difference in rating between heterosexual group and group of other sexualities.

Question 3: Results

A summary table of the coefficients of the linear regression model.

```
##                          Estimate Std. Error   t value   Pr(>|t|)
## (Intercept)             47.485701  0.1903805 249.42517 0.0000e+00
## sexuality_groupothers    8.098851  0.5987245  13.52684 1.3935e-41
```

According to the summary table, β_0 is about 47.49, which shows the estimated mean rating for liberal party from the heterosexual group, and the estimated mean rating of liberal party for other sexualities ($\beta_0 + \beta_1$) will be 55.59. Furthermore, the **p-value** is very small (1.3935×10^{-41}) in this model, so we have strong evidence against the null hypothesis that the rating of Liberal party is the same for heterosexual voters and other sexualities. **Therefore, there is an association between sexuality of voters and the rating of the Liberal party.**

Conclusions & Limitations

Based on the three questions that we investigated, here are the **conclusions** and **suggestions** that we would like to inform:

Question 1:

The result implies that when making policies, Liberal party should be aware of the welfare of those who are poorly educated and provide them the equal opportunity of schooling.
Limitation: There are 2646 people who chose not to give their education information, which may lead to a slightly skewed result. Also, people who attended "some university" have relatively higher party ratings as well – including them in the study may influence the result.

Question 2:

Through our analysis, we found that Liberal Party's rating did not differ between working people and retired people, indicating that the current policy is relatively fair to both groups of people.

Conclusions & Limitations

Question 2:

(Cont.) If a new policy is to be implemented, it is better to take care of both groups of people, otherwise it may lead to dissatisfaction.
Limitation: We cannot tell whether a person is working or not accurately just based on the age, for example, individuals may still be willing to work even they are above 65 years old.

Question 3:

Liberal Party gains higher rating from sexuality groups other than heterosexual. This may reveal that Liberal party have great LGBT policy to appeal voters who are homosexual, bisexual, or belong to other minor sexuality groups, and they should keep it in order to gain more votes.
Limitation: Since a number of participants have chosen "prefer not to answer" and some of the participants' sexuality is not known, the result will be different if the information is considered.

Conclusions & Limitations

Overall, we investigated about the **rating of Liberal party** from three social perspectives. A general limitation is that the party rating may not totally indicate the vote choice of individuals. For example, individuals who give a high rating to the Liberal Party may not vote for it eventually. Therefore, we still need to investigate about vote choices further.

However, the ratings do indicate **preferences**, and differences in rating from different groups surely help the party to keep or adjust their policies, in order to win more votes from more groups.