Question 1

Jyoti Kumari

Contents

1 Are Democratic voters older or younger than Republican voters in 2020?

1.1 Importance and Context

The primary research question we're looking at in this report is around the age of democratic and republican voters. Age is an important demographic variable for election cycles. By focusing on the age, we can analyze our voters for the different political parties. For instance, younger voters can sometimes be less likely to vote due to the friction of registration to vote, busy life, geographic mobility, and sometimes just not having developed the habit of voting. With this insight, political campaigns can focus their initiatives on the diverse age groups.

1.2 Description of Data

The ANES 2020 Time Series Study Preliminary Release: Pre-Election Data is being made available in two statistical file formats: SPSS (.sav) and Stata (.dta). Version used ANES2020TimeSeries_20210211

Several variables were used to analyze the research questions such as V201018, V201507x, V201028, V201038, V201008, V201048, V201066, V201101, V201104, V201075x, V201076x, V201077x, and V201078x. These columns contained information to define voters, identify values for age, vote registration status, and political party.

From the original ANES dataset, a series of dataframes were created to extract the columns that were relevant to our research question and for plotting purposes. A temporary dataframe was used to clean up the data (eliminating data from other political parties and non voters). Voters were defined as people that were registered to vote, if they either voted for a presidential, house of representatives, senate, governor, also if they vote for President in the 2012 or 2016 election. In addition, the summary data of vote/intent/preference was considered.

For the age variable, records of people that refused to respond were eliminated since the age data is crucial to answer our research question. The age population of 80 of greater than 80 was reported as 80 value, even though it causes the data to pick at 80, data was preserved since our research question is related to the age variable. Political party was limited to the two political parties of interest.

After the data was filtered and cleaned, a graph was created to analyze the data distribution. See Figure 1. The Age distribution for both political parties is not normally distributed, and the data is negative skewed for both of the parties. As we mentioned before, the age population of 80 of greater than 80 was reported as 80 value causing the data to have a spike at the 80 age value.

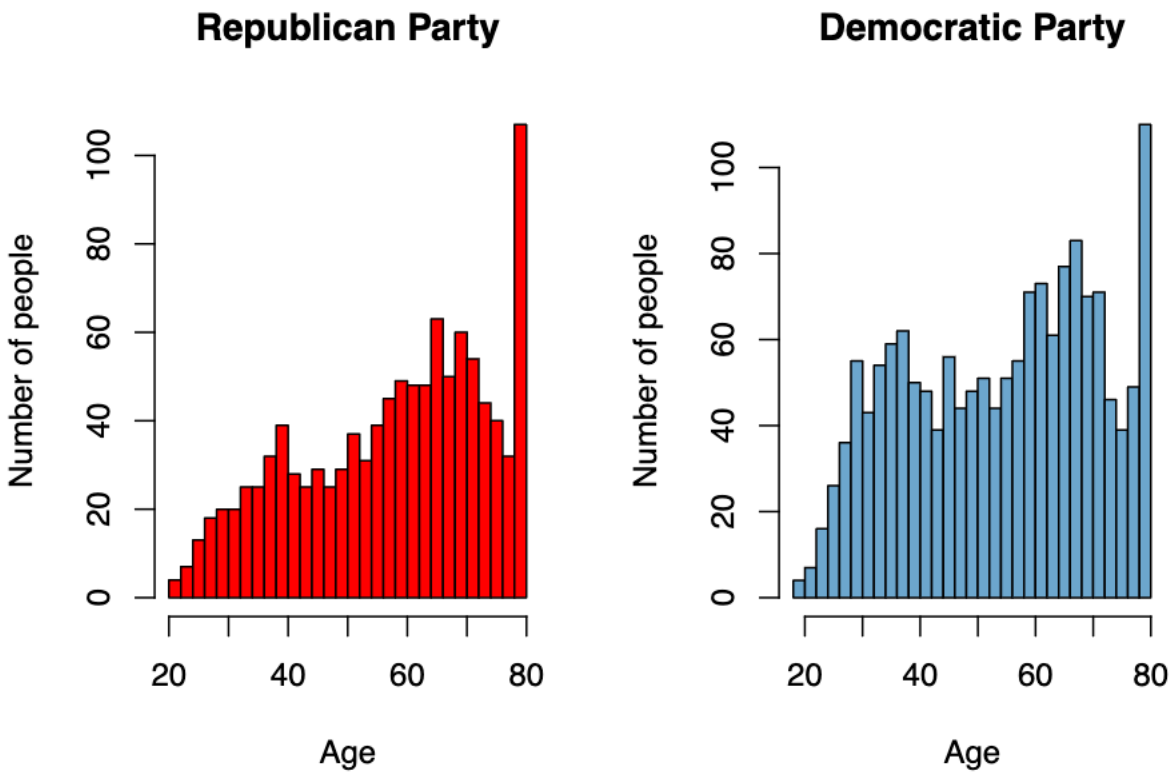

Figure 1: Age Distribution by Political Party

1.3 Most appropriate test

Breaking down the data per Political Party Affiliation (Republicans or Democrats), we will get two groups of voters. The response variable is Age. Since age is a metric variable and we have two distinct groups, we can use employ the 2Sample Welch's T Test here.

Looking at the assumptions for the test: IID - The ANES 2020 uses respondents from a fresh cross-section of respondents, and a collaboration to interview a subset of respondents from the General Social Survey (GSS). This data collection process also rewards individuals for filling out surveys. There is a possibility that this introduces dependencies. For example, participants may tell friends or family members about this survey, resulting in a cluster of individuals that give similar responses. Nevertheless, since the data collection process claims to have millions of users, which suggests that links between individuals should be rare. Metric - Age is a metric variable. A metric variable needs to have intervals to be equivalent. Here, age will fulfill that requirement. Normally Distributed - We can see in Figure 1 that the data is skewed and has a spike at 80. However, with a high sample size, CLT will kick in and normality will be fulfilled.

Hypothesis:

$H_0 : Mu_{(Democrats)} = Mu_{(Republicans)}$

$H_A : Mu_{(Democrats)} ! = Mu_{(Republicans)}$

```
t.test(df_democrats_republican$Age ~ df_democrats_republican$PoliticalParty)
```

```
##
##  Welch Two Sample t-test
##
## data:  df_democrats_republican$Age by df_democrats_republican$PoliticalParty
## t = -5.4954, df = 2394.9, p-value = 4.31e-08
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -4.686915 -2.221686
## sample estimates:
## mean in group 1 mean in group 2
##        54.57885        58.03315
```

1.4 Test, results and interpretation

We reject the null hypothesis that the two means are the same at the 95% confidence level. We have a highly significant test. We can reject the null hypothesis that the mean age of Democratic voters is the same as that of Republican voters in 2020.

In this test we are comparing the two group means and the test determines that they are not significantly different. The mean of group 1: Democrats is 54.57885 while the mean of the group 2:Republicans. The difference in means is 3.4553 years. Practically the difference in mean is quite small.

Going back to the research question, Are Democratic voters older or younger than Republican voters in 2020? We were able to determine that the age means at the 95% confidence level for Democratic voters and Republican voters it's not the same, however the difference in means it is not significant.

1.5 References

American National Election Studies. 2021. ANES 2020 Time Series Study Preliminary Release: Pre-Election Data [dataset and documentation]. February 11, 2021 version. www.electionstudies.org