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

Catherine Tsai, Mai La

Importance and Context

Democrats are commonly referred to as "progressive or liberals" while Republicans are commonly referred to as "conservatives." Pew Research shows that political associations shift further from liberal towards conservative view points as age increases.

Age is a factor in determining life stages. Example, a voter in their 20's will care very differently about Medicare coverage versus voters in the 80's will care very differently about student loan forgiveness and the policies each party will champion. Understanding voter age is important as it will give insight and inform decisions about what types of issues that will motivate voters to show up on Election day. Each party must strategically apply efforts to woo their target audience to impact the final outcome on game day (i.e. Election day).

Population growth and changes can also determine the popularity of a party if the policies resonate with one group vs. another. Example, if population is shrinking, the overall population will age in futures years and would turn more supporters towards a party that attracts older populations. To understand demographic makeup of voters age and impact on politics, the study seeks to understand the following question: Are Democratic voters older or younger than Republican voters in 2020?

Description of Data

In order to answer the question, we will use data from the 2020 American National Election Studies (ANES). We define a voter using several survey questions pre and post election and exclude any records that indicate the person is a non voter in the 2020 Election (i.e. not registered to vote, do not intend to register to vote, did not vote, etc.).

To determine if a voter is identified as a Democrat or Republican, we identified party affiliation from pre election survey data. In order to have the most complete view of Political party identification, we used Political party variable with the most number of records for Republican and Democrats and classified left leaning as Democrats while right leaning as Republicans. Responses that could not be classified into Democrat or Republican (such as Independent and or "do not know" were excluded from our analysis).

The numerical age was provided in the analysis for those responses between the age of 18-79. A second grouping of data for responses "80 or older" was excluded from our data set as exact age was not provided. In order to more accurately determine expected age of both parties, we avoid estimating or imputing data without having a relative expectation of accuracy.

Segmenting for responses related to the age and party using the criteria above we end with 3,499 responses for Democrat and 2,991 responses for Republican in our study (6,490 total responses).

Table 1: Sample Size of Democrats & Republicans Used in Study

Party	Number of Samples
Democrat	3499
Republican	2991

| Party	Number of Samples
Total	6490

Looking at the age distribution (top graph) of Democrats and Republicans in our study, it seems that Republicans skews to the right (older). Democrats have many more records in age 20-40 vs. Republicans. Democrats show 2 peaks for at age 40 and 60. Normalizing the data due to the higher number of Democrat records in our data, the skew towards older demographic is more pronounced (bottom graph). Overall, there is not a very notable difference in the distribution between the two parties.

Age Distribution Among Democrats and Republicans
Democrats have more records in younger age, Republican skews older
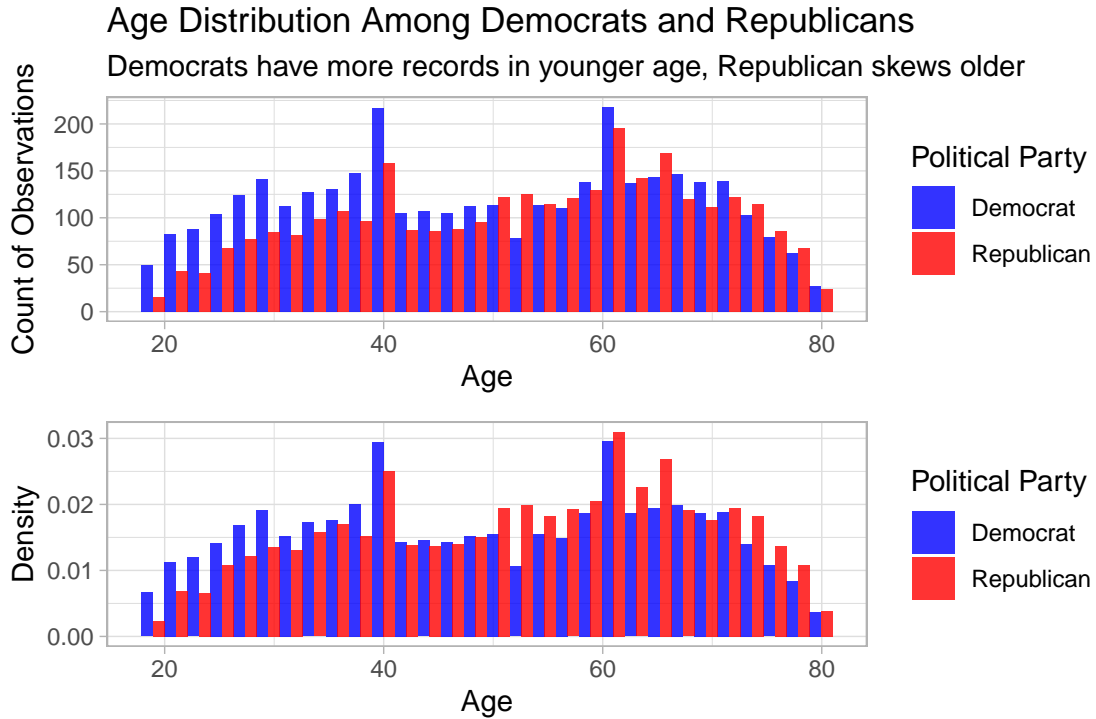
In the boxplot below, we further summarize the interquartile age range of both parties. The lower quartile and median age for Republicans is higher than Democrats (older). Republicans have a higher concentration of records between the 50-75th percentile relative to Democrats where the quartiles are more evenly distributed between 25-50th percentile and 50th-75th. Democrats seem to have better voter coverage across the different age ranges.

Boxplot of Age Distribution Among Democrats and Republicans
Democrats age is more evenly distributed versus Republicans
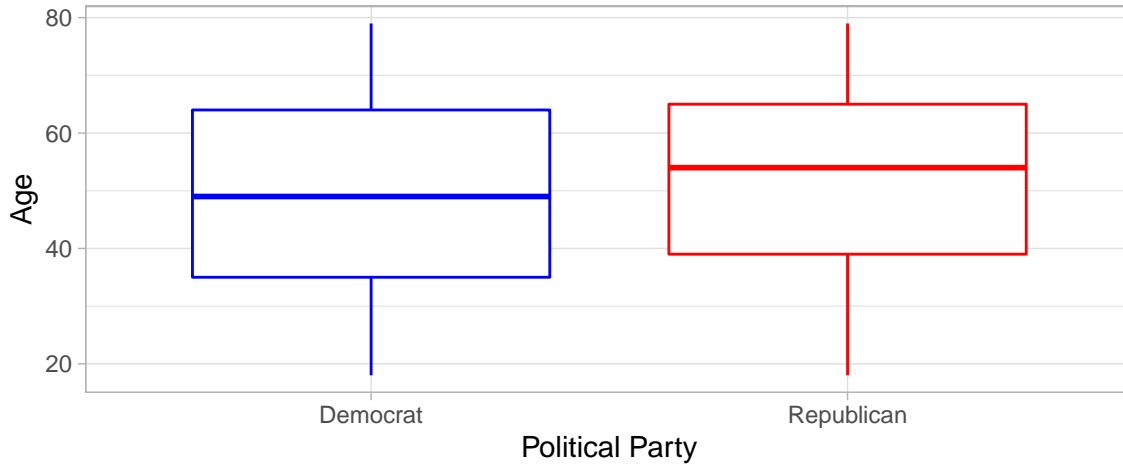
Most appropriate test

Based on our stated question and study data, the most appropriate statistical test to conduct is a two sample, two tailed Welch t-test.

The study compares different groups: Republican and Democrats and they are independent of each other (not paired - based on different sample sizes of within our data). We will use a two tailed test in order to understand any difference in age between the groups - either older or younger relative to each group. We do not want to ignore any effect that would be statistically notable in one direction (in favor of another).

Assumptions:

- Data is of a metric scale
  - Age is a measurable and on an interval sale. Therefore it is metric in nature.
  - One can compare and say a person of age 50 is older than age 40.
- Independent and Identically Distributed (IID)
  - The methodology of the data collection seemed to be more strenuous and mixed (Web, Phone, Zoom) in previous years due to COVID as well with several attempts needed and higher bounties placed on non responses.
    * This could lead to potentially skewed data with mixed respondent motivation based on 1) free time to participate in survey and 2) motivation/need for survey response fees. Certain demographics/age groups (such as retired folks) could be more prone to respond and bias the results.
  - The target population is a cross section of 231 million US citizens age 18 or order
  - 2020 ANES sample is based on a random draw from US across 50 states with equal probability
    * The sampling methodology design accounts for sample bias across states.
- No major deviations from normality
  - From the histogram plot of age and Political party, we do not see any major deviations from normality.
  - Given the sample size of our data (2.5K+ records for each party) we can use the central limit theorem to assume sample mean will converge to a normal distribution

Given the data and the research question, we form our hypothesis:

Null Hypothesis: The average age of Democratic voters *is equal* to the average age of Republican voters

$$H_o : \mu_1 = \mu_2$$

Alternative Hypothesis: The average age of Democratic voters *is not equal* to the average age of Republican voters

$$H_a : \mu_1 \neq \mu_2$$

Rejection Criteria:

$$p - value < 0.05$$

Using a two tailed t-test with p-value of 0.05, we proceed to conduct the t-test.

The conclusion from the experiment should not be interpreted as a casual inference but rather observational. The experiment can provide information regarding expected ages from the two parties based on the data but cannot be interpreted as age causing an effect on political association. The results of the test cannot provide insight on the factors that make people associate with one party or another. The test is merely an observation on the differences average expected age of Democrats and Republicans in the 2020 election.

Test, results and interpretation

```
## Perform t-test between two groups
t.test(age~party, data = age_party)
```

```
##
##  Welch Two Sample t-test
##
## data:  age by party
## t = -7.2894, df = 6425.8, p-value = 3.486e-13
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.700592 -2.132033
## sample estimates:
##    mean in group Democrat mean in group Republican
##                  49.36553                 52.28185
```

From the t-test conducted, there is statistically significant evidence (p-value below 0.05) to suggest that the average age of a Democratic voters is not equal to that of a Republican voters. We would reject the null hypothesis from a statistical viewpoint based on our test criteria above.

From a practical viewpoint, the difference of 3 years is not that significant in that the viewpoints, political issues of interest, and lifestage perspective is not significantly different among adults 49 versus 52 years of age. The calculated Cohen's D at -0.18 is also at negligible level.

Beyond the expected voter age provided by the test, there is evidence to show (as seen in the Boxplot of Age Distribution Among Democrats and Republicans) the makeup of Democratic voters in our sample has more records in the younger age range. These types of insights can help inform party decisions. To play on their existing advantage, Democrats should try to "market" policies more towards younger people if they want to attract more voters and support. Alternatively, Republicans should try aim their policies to ones that appeal to an older crowd.

Additionally, there is more opportunity for both parties more voter by appealing to younger people. The 18-29 voting eligible population has the lowest voting ratio. This highly coveted demographic is difficult but not impossible to attract. Younger voters may be harder to engage, but older voters are reliable to show up and vote (as seen through our sample data). Understanding the dynamics of the voter base is important in shaping the policies and political strategy of each party.