Lab 1: Question 1

Group 2: Andres de la Rosa, Elizabeth Khan, Carlos Moreno

Contents

Importance and Context

Question of interest: **Are Democratic voters older or younger than Republican voters in 2020?**

The demographic profile of voters in the US has changed over the last decades. As the US population gets older, so does the profile of US voters. A 2019 Pew Research study ("The Changing composition of the Electoral and Partisan Coalitions") has found this age trend to be more common among Republican voters than Democrats. The 2019 report stated that most registered voters in 2019 (52%) were ages 50 and older, representing a minimal change from 2012 (51%). However, it is much higher than in 2004 (44%). In addition, in 2019, while 56% of Republican voters were ages 50 and older, a smaller share of Democratic voters was in that age group (50%).

The ANES 2020 Time Series Study (preliminary results) seems to support this demographic trend. As presented in Table 1, while 57% of overall voters are 50 years of age and older, for the Democrats only 53.1% of voters were 50 years of age and older, versus 61.8% for republicans.

Table 1: Age Split by Party Affiliation

Age_Group	*Party_Affiliation*		***Total***
	Democrat	Republican	
(0,30]	440 14.4 %	239 9.3 %	679 12.1 %
(30,50]	996 32.5 %	739 28.9 %	1735 30.9 %
(50,80]	1626 53.1 %	1579 61.8 %	3205 57 %
Total	3062 100 %	2557 100 %	5619 100 %

χ^2=53.303 · df=2 · Cramer's V=0.097 · p=0.000

A change in age composition for the US political parties could have profound implications for how politicians run their campaigns. If confirmed, age could be driving the shift of public opinion for important social issues, which could change how political parties interact with constituents and potential voters, including the selection of social media technologies and the development of more targeted messages that are pertinent to this generation.

Another implication is that age is a recognized demographic variable impacting voter engagement. It has been established that older voters are more likely to participate in the democratic process versus younger voter. Thus, the proposed research question has profound implications for political parties and the policies that will govern the US in the decades to come.

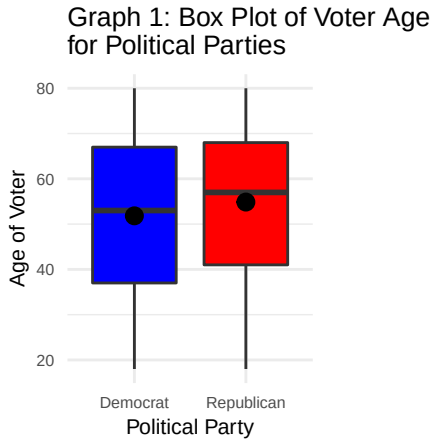
Description of Data

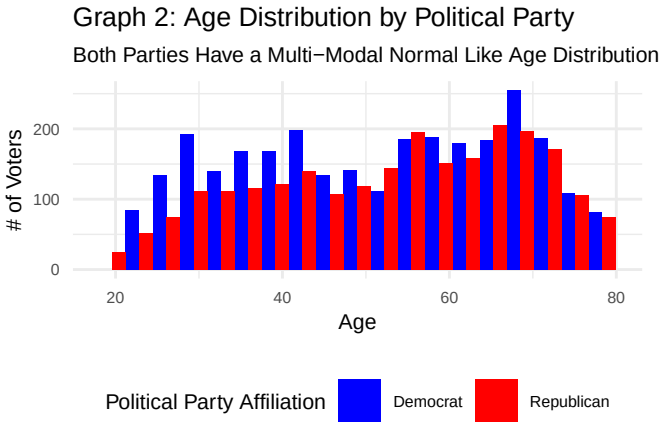
To answer the question of interest, we will use the preliminary results from American National Election Studies (ANES) 2020 Time Series Study. This study combines pre-Election and post-Election survey results. Participants were reached via internet, phone, and video. Pre-election interviews began in August 2020 and continued until Election Day (Tuesday, November 3rd). Post-election interviews began soon after the election and continued through the end of December 2020.

To answer the question, we focused on participants identifying themselves as Democrats (3063 voters) or Republican (2558 voters) in the post-election survey using the variable V202443 ("Which party does Respondent feel closest to?") renamed as "Party_Affiliation". For the age of the voters, we used the variable V201507x renamed as "Voter_Age", which includes the Respondent Age.

Graph 1 includes a Box-Plot for the age of Democrat and Republican voters. The ages for both parties ranges from 18 to 80 years of age, where 80 was assigned to any voter 80 years or older. This graph also shows a slightly younger age distribution for Democrats when compared to republicans. The black dot in the graph identify the average age for each group.

In this dataset, Democrats have an average age of 51.8 years with a standard deviation of 17.3 years, while Republican voters have an average age of 54.9 years with a standard deviation of 16.5. Thus, the difference in average years between Democrats and Republicans is of -3.1 years. Graph 2 shows that both of the age distributions for Democrats and Republicans are multi-modal with a slight skew to the left.





Most appropriate test

To test the hypothesis, we propose a t-test to compare two unpaired means. There are four important conditions for the proposed unpaired t-test:

1. Unpaired data: the grouping variable is party affiliation (Democrat or Republican), and as each respondent can only be affiliated with a single party, we consider this an unpaired data.
2. The variable of interest must be metric: the variable of interest is *Age of Voters*, which is a metric variable where basic arithmetic operations such as average are possible.
3. I.I.D: The design of the study suggests the data was generated through an **independent and identically distributed (i.i.d.)** sampling process. Participants in this survey were randomly drawn from the USPS delivery sequence file. The USPS delivery sequence file contains all residential addresses in the 50 states across the U.S. and all addresses had an equal chance of being selected.
4. Normally Distributed: *Graph (2)* shows that the age distribution, for both Democrats and Republicans, is multi-modal, slightly skewed to the left as 57% of voter in the sample are 50 years of age and older (53.1% for Democrats, 61.8% for Republicans). In addition, the large sample size of 3063 Democrats and 2558 Republicans, supports a normal like distribution based on the Central Limit Theorem (CLT).

For the t-test, the null and alternative hypothesis are:

- **Null Hypothesis (Ho)** - There is no difference between the average age of Democrat Voters when compared to average age of Republican Voters.
- **Alternative Hypothesis (Ha)** - The average age of Democrat Voters is different from the average age of Republican Voters.

In short, this report aims to find out if there are significant differences in the average age of Democrat voters when compared with average age of Republican voters. These hypotheses can be expressed in the following way:

$$H_0 : \mu_D - \mu_R = 0$$
$$H_a : \mu_D - \mu_R \neq 0$$
$$Where : \mu_D = Average\ Age\ of\ Democrats, and\ \mu_R = Average\ Age\ of\ Republicans$$

We will use a two-tailed test, as this is the strictest test which would help identify if there are significant differences (or not) in the average age of the two groups (Democrats and Republicans).

We will use p-value as the rejection criteria, with an alpha of 0.05. Thus, if the t-test p-value is less than 0.05, we would reject the null hypothesis (Ho), otherwise, we will fail to reject it.

Test, results and interpretation

With a p-value of 1,69942594e-11 which is less than alpha (0.05), we reject the null hypothesis that the average age of Democrat Voters is equal to the Average Age of Republican Voters. Thus, there is evidence suggesting that their average age is different. As the difference in average age (-3.1 years) is negative (Confidence Interval (CI): -3.9, -2.2), it suggests that the average age of Democrats is lower than average age of Republicans.

Though, the difference in average age of -3.1 years between Democrats (51.8 years) and Republicans (54.9 years) is significant, it seems relatively small. In practical terms, within the span of a person's life, -3.1 years is a relatively small difference. The average age is hiding an important trend which can only be observed by tracking the change in age structure through time, such as - How age group distribution is changing through time? Which age groups are taking on more relevance?

As discussed at the beginning of this report, a 2019 Pew Research study found that though US voters have aged as the US population ages, this trend has been larger among the Republican voters when compared to Democrats Voters. The findings in this report seems to align with this trend, as the data suggests that the average age of Democrats is lower than average age of Republicans.

The change in the voters age structure is expected to influence the way politicians run their campaigns. Politicians will need to adjust the way they reach potential voters, and modify the type of messages they use, as younger voter will continue to drive the shift in public opinion in some important social issues.

Test Limitations

One limitation of the t-test is that it focuses in comparing overall averages. While the average is a convenient metric, it hides key information such as the actual age structure for each party. It would be important to identify the age groups taking on more relevance, and how this has changed through time. As it can be seen in Table 1, it seems that the group (30 to 50] and (0 to 20] years of age are becoming important age groups for Democrats.

References:

(1) Pew Research Center. (2020, Jun 2). "The changing composition of the electorate and partisan coalitions". Pew Research Center - U.S. Politics & Policy. https://www.pewresearch.org/politics/2020/06/02/the-changing-composition-of-the-electorate-and-partisan-coalitions/

(2) American National Election Studies (ANES) - 2020 Time Series Study (Preliminary Results). 2021, Mar 24. https://electionstudies.org/data-center/2020-time-series-study/