Lab 1: Question 1

Ian Dela Cruz, Steve Hall, Fengjiao Sun

Contents

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

Importance and Context

The 2020 election was arguably the most important in modern history as the political divide was driven wider by the Covid-19 crisis, economic fallout, and social unrest. A major factor typically used to explain this widening ideological gap is age. A Pew Research Center [1] study found that an increasing number of Gen X have identified as liberal and Democrat while Boomers have turned more conservative and Republican. Furthermore, a Chicago Booth Review article [2] suggested that this has been the case for generations as older Americans, who now identify as conservative, were liberal in their 20s and 30s. With an aging population, how will these changing demographics influence campaign strategy and politics for decades to come? There are several questions we could aim to answer but first, we will start by simply answering:

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

Description of Data

We will address this question using data from the American National Election Studies (ANES) time series study. The 2020 ANES survey used a contact-less, mixed-mode design that was created in response to challenges related to the COVID-19 pandemic. For instance, the face-to-face mode was dropped. Instead, a sequential mixed-mode design was implemented that included self-administered online surveys, live video interviews conducted online, and telephone interviews. Pre-election interviews were conducted from August 2020 until Election Day (Tuesday, November 3). The study has a total of 8,280 pre-election interviews. See code book for full details on collection methodology [3].

To answer our research question, we will only need to use two variables from the survey: party of registration (V201018) and respondent age (V201507x). Since we are only concerned with the age difference between Democrats and Republicans, we will exclude the other party categories such as 'independent' and 'other'. Similarly, we will also exclude records in which the respondent did not provide an age.

After sub-setting the original data set, there were 1808 registered Democrats and 1266 registered Republicans surveyed with appropriate ages.

It is also important to note that the age variable has been transformed by representing all ages over 80 with 80. This truncation of the respondent age variable forces us to make a couple assumptions that we will discuss later.

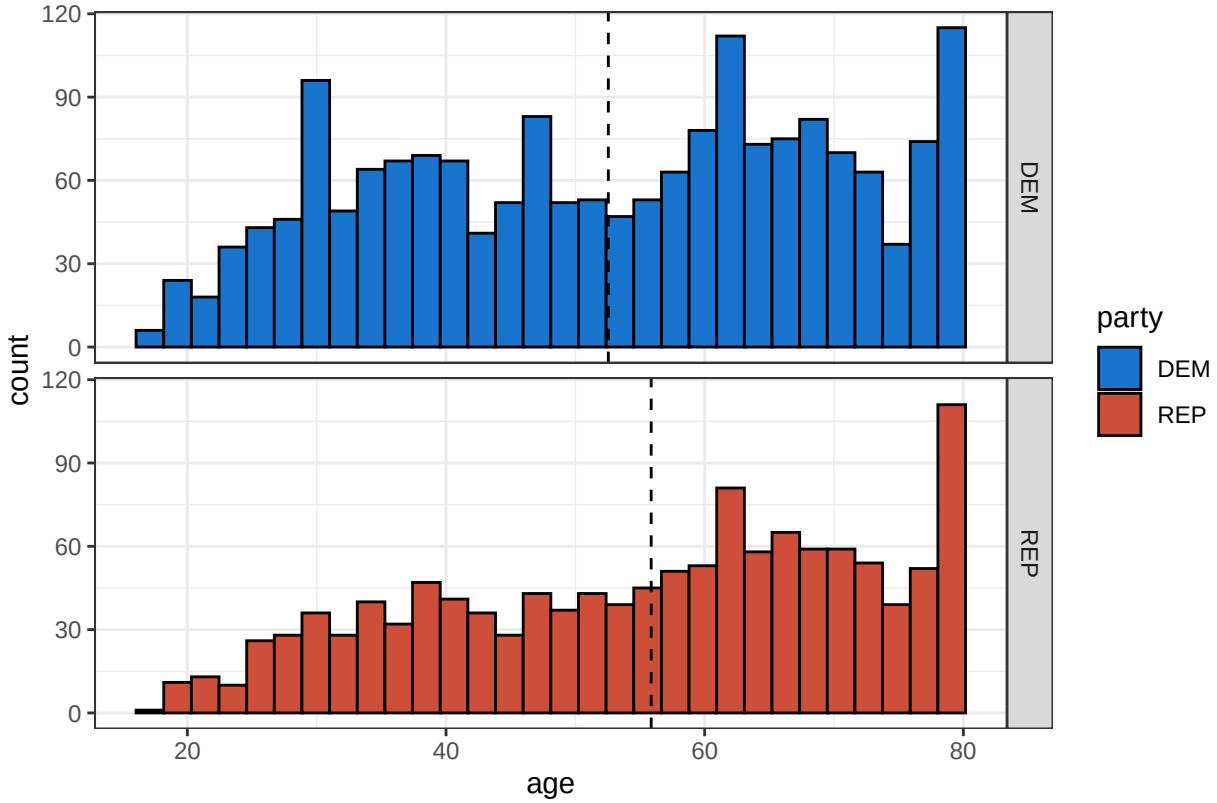
At a quick glance, you can see in figure 1 that Democrats skew slightly younger. This is also evident by the slightly lower average age, which is represented by the dashed line. Lastly, a similar number, but lower percentage of Democrats are 80 years of age or older. To be precise, 5.59% of Democrats were 80 or older while 8.21% Republicans fall in that category.

[1] https://www.pewresearch.org/fact-tank/2017/03/20/a-wider-partisan-and-ideological-gap-between-younger-older-generations/

[2] https://review.chicagobooth.edu/economics/2020/article/there-are-two-americas-and-age-divider

[3] https://electionstudies.org/data-center/2020-time-series-study/

Figure 1: Age Distribution of Democrats and Republicans in 2020



Most appropriate test

Typically, we would consider age as a nominal variable on a metric scale if it were not transformed. In this case, however, since the value of 80 represents all ages greater than 80, we argue that this truncated age distribution is now on an ordinal scale. In other words, we can no longer subtract 79 from 80 and say that is one year of age difference. All we know is that 80 is greater than 79. For this reason, we chose to run the Wilcoxon Rank Sum Test to compare the age of Democrats to the age of Republicans instead of a two sample t-test.

The **Wilcoxon Rank-Sum Test** makes two important assumptions:

- **Ordinal scale**: The variable of interest is age and ranges from 18 to 80 (truncated). As discussed above, since the age variable has been transformed we assume it is ordinal in nature.
- **IID data**: Even though there may be some correlation of age between the population in terms of who is surveyed across Democrats and Republicans, we assume that the samples are independent. Furthermore, the samples are drawn from the identical distribution. Therefore, we can assume that the distributions are identical and independently distributed.

The null hypothesis that we are testing, $P(DEM > REP) = P(DEM < REP)$, is that the probability that a draw from DEM ranks higher than a draw from REP is the same as the probability that a draw from REP ranks higher than a draw from DEM.

If this test were to **reject the null hypothesis**, we would conclude that there is a statistically significant difference between the ages of Democrat and Republican registered voters in 2020. If the test were to **fail to reject the null hypothesis** then we would conclude that there is no difference in age between the two major political parties.

Test, results and interpretation

Using the unpaired two-sample Wilcoxon test in R, we can find the test statistic and p-value for the sample, shown below.

```
wilcox.test(age ~ party, data = df_clean)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  age by party
## W = 1018182, p-value = 1.831e-07
## alternative hypothesis: true location shift is not equal to 0
```

Given the extremely small p-value of 1.831×10^{-7}, we **reject the null hypotheses**, in favor of the alternative that the age of Democrats is different than that of Republicans.

Although the test is positive and we reject the null, the difference in age between Democrats and Republicans is only -3.3 years. A difference of this size given the range of ages is a small effect. Furthermore, Cohen's D, a common measure for effect size, is only -0.18. This also suggests a small effect, or difference between the age of Democrats and age of Republicans.

While there is a statistically significant difference, the practical significance is less compelling as an approximately three year average age difference is not meaningful in real life. We note that this survey was completed before the actual election, but from this data alone, it would be difficult to persuade most people that age was a major contributing factor to the election results. Finally, this raises the question: does a small age difference between Democrats and Republicans drive such polarizing views on the economy and social issues? Or are there other factors that have widened the ideological gap between the left and the right?