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

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

The 2020 election was arguably the most important in modern history as the political divide was driven even 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 study found that an increasing number of Gen X has identified as liberal and Democrat while Boomers have turned more conservative and Republican. A Chicago Booth Review article 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 a population that is expected to age over the next few decades, how will that affect politics and policy? Will Democrats be able to retain Gen X and Millenial loyalty as they age?

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

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

Data Description

We will address this question using data from the American National Election Studies (ANES) time series study. The ANES 2020 Time Series Study is a continuation of the series of election studies conducted since 1948 to support analysis of public opinion and voting behavior in U.S. presidential elections. This year's study features re-interviews with 2016 ANES respondents, a freshly drawn cross-sectional sample, and post-election surveys with respondents from the General Social Survey. The study has a total of 8,280 pre-election interviews.

To answer this question we will only need to use two responses, or variables from the survey: party of registration and respondent age. Since we are only concerned with the age difference between Democrats and Republicans, we will exclude the other party categories like 'independent' or 'other'. Similarly, we will also exclude records in which the respondent did not provide an age.

After sub-setting, 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. By simply looking at the plot, 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.

Figure 1: Age Distribution of Democrats and Republicans in 2020
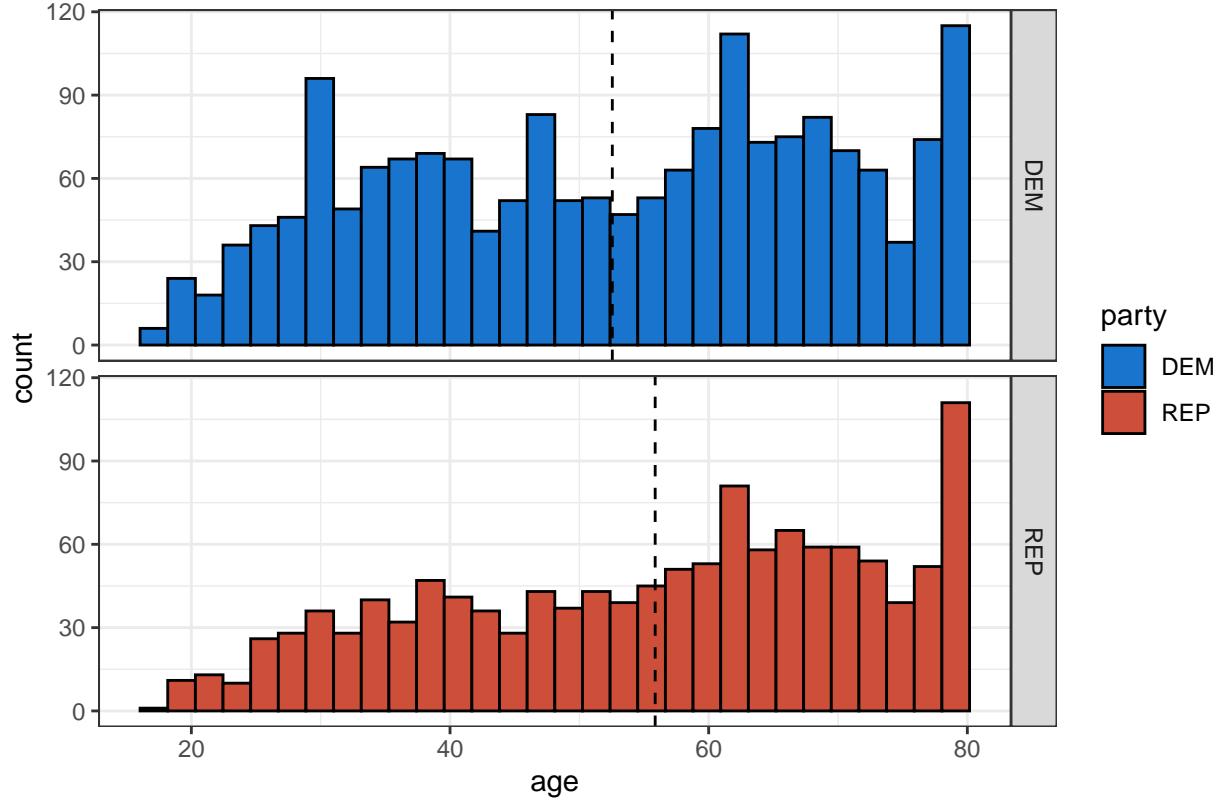
Most appropriate test

Typically we would consider age as a nominal variable on a metric scale if it weren't 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 that reason, we chose to run the Wilcoxon Rank Sum Test to compare the age of Democrats to the age of Republicans. Specifically, we will run the Hypothesis of Comparisons version since the transformed age variable is on an ordinal scale. In addition to the assumption that age is on an ordinal scale, we are required to assume that the data is IID. The age variable is recorded from each respondent and completely independent from the other respondents of the survey.

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.

Now, it is also reasonable to assume that the transformed distribution is approximately the same as the raw underlying distribution since the transformation is applied to both parties and it is relatively infrequent (less than 10% of all observations for each party). In this case and given its robustness to non-normality and the relatively large sample sizes (n = 3074), we believe the two sample t-test is also appropriate, but less conservative as the Wilcoxon Rank Sum test. For the two sample t-test, the hypotheses are as follows:

$$H_0 : \mu_D = \mu_R$$
$$H_1 : \mu_D \neq \mu_R$$

3

Test, results and interpretation

For both tests we **reject the null hypotheses**, in favor of the alternative that the age of Democrats is different than that of Republicans. The p-values were extremely small and similar at 1.831×10^{-7} and 1.017×10^{-7} for the Wilcoxon Rank Sum and two sample t-test, respectively.

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. It would be difficult to convince anyone in either party to change their campaign messaging on this information alone.