Question 3

Jyoti Kumari

Contents

1 Governer's Rating and Covid

1.1 Importance and Context

The primary research question being looked at in this report is "Are survey respondents who have had someone in their home infected by COVID-19 more likely to disapprove of the way their governor is handling the pandemic?" The pandemic of COVID-19 has affected the vast population diversely, gravely and left the public with strong opinions about the government and authority. This research question will help us understand how the public is feeling about their leadership, that is, the governor's actions and policies around the pandemic depending upon whether they have had someone near and dear affected by Covid or not.

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. This is the data that we'll be using.

The features being used for this research question are the following:

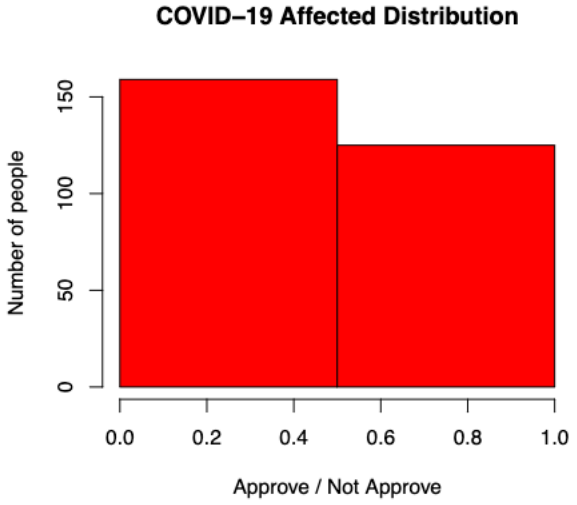
V201624 - PRE: ANYONE IN HOUSEHOLD TESTED POS FOR COVID-19.

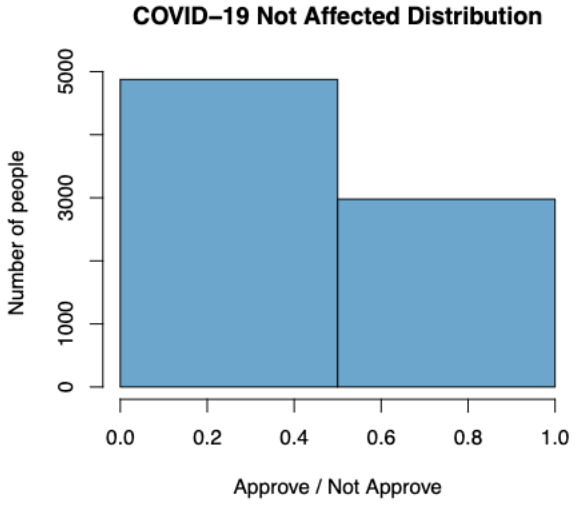
This one provides us with the information if anyone in the respondent's household has contracted with Covid-19. The set of values in this feature are filtered to only 1 (Someone in my household tested positive) and 2(No one tested positive). This feature is going to be used as a grouping variable to create two distinct groups - the respondents who had someone in their household test positive vs the respondents who didnt have anyone in their household test positive. There were other values in the feature such as -9(refused) as well but it has been filtered to only 1 and 2.

V201145 - APPROVE OR DISAPPROVE R'S GOVERNOR HANDLING COVID-19

This feature is going to provide the respondent's opinions about the governor's handling of the pandemic. The values are filtered to 1 (Approve) and 2 (Disapprove). The other values have been filtered out like refused and NA. This is the primary variable which will help us analyze if the the respondent is more likely to disapprove of their governor's handling of the pandemic.

For each of these variables, they have been changed from 1 and 2 to 0 and 1. The graph below shows the distribution of governor approval vs disapproval for the two groups of someone in the household affected by covid vs not affected by covid.





1.3 Most appropriate test

The most appropriate test for this research question is Welch's 2 Sample T Test. Using "V201624 - PRE: ANYONE IN HOUSEHOLD TESTED POS FOR COVID-19" as the grouping variable, there are two sample created for "V201145 - APPROVE OR DISAPPROVE R'S GOVERNOR HANDLING COVID-19". Thus with two samples and the governor's rating as a metric variable, Welch's 2 Sample T Test can be used. The test will be an unpaired one.

Mean rating of the governor's handling of the pandemic is calculated for each group in the test.

The primary assumptions for Welch's 2 Sample T Test are:

1. IID - As a modern survey, the GSS looks for each respondent beginning from scratch to avoid dependencies. One might worry that some people are more likely than others to respond. Moreover, people in different states can have varying opinions on the leadership by the governor. And since different states are affected differently by the pandemic, that can have an effect on the responses as well. However, this is more of an issue around generalizability. It doesn't have an impact on IID. It just means that the entire population might not be represented holistically. However, the test still remains valid. Since there is only a finite number of Americans, it is impossible to have perfect independence. Once a person is drawn, they cannot be drawn again, so the distribution for the next draw has changed. However, given there are a substantial number of Americans, the change is likely to be very small, so we can safely ignore these finite-sample effects (in general something to worry about with much smaller populations). On the whole, there are issues to worry about, but modern surveys like the GSS do a pretty good job of making the iid assumption valid.
2. Metric - The outcome variable of governor's rating has two values of 0 and 1. Metric variables require intervals to be equivalent. Here, as a binary variable, there is only one interval - from 0 to 1.
3. Normality - Within each group of covid affected vs covid non affected household, we can see that the bimodal distribution is not hugely imbalanced. Moreover, a bimodal distribution works for CLT to kick in. Since there is a large sample size here, CLT will kick in and the condition of normality is fulfilled. You can see this in the Figure 1 above.

With the requirements for the test verified, we're going to proceed with the hypothesis testing.

Our null hypothesis is: $H_0 : mu_{\text{(Households affected by covid)}} = mu_{\text{(Households not affected by covid)}}$

Alternative hypothesis is: $H_A : mu_{\text{(Households affected by covid)}} \neq mu_{\text{(Households not affected by covid)}}$

Here, the mu is being calculated for the governor's rating wherein 0=approve and 1=disapprove.

Conducting the test:

```
t.test(filtered_data$governer_rating ~ filtered_data$covid_affected)
```

```
##
##  Welch Two Sample t-test
##
## data:  filtered_data$governer_rating by filtered_data$covid_affected
## t = 2.0316, df = 302.81, p-value = 0.04307
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.001912944 0.120029047
## sample estimates:
## mean in group 0 mean in group 1
##       0.4401408       0.3791698
```

1.4 Test, results and interpretation

The p-value of the test with a confidence interval of 95% is 0.04307, less than 0.05 and thus statistically significant. We can reject the null hypothesis.

However, having looked at the statistical significance, we are now going to look at the practical significance. The mean of governor rating of the group with someone affected by covid in their household is 0.44 while the mean of the group with no one in their household affected by covid is 0.38. Calculating the difference in means, 0.44-0.38, we see that it is only 0.06. Thus those who have someone in their household test positive are 0.06 more likely to disapprove of their governor than those who had no one affected by covid in their household. The difference is not significant practically, even though it is statistically. The difference in rating is so small practically that it doesn't suggest anything strongly.

Therefore, closing in for the research question of "Are survey respondents who have had someone in their home infected by COVID-19 more likely to disapprove of the way their governor is handling the pandemic?" - No, they are not more likely!

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