Lab 1 Question 3: 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?

Yao Chen, Jenny Conde, Satheesh Joseph, Paco Valdez, Yi Zhang

Contents

1 Importance and Context

The COVID-19 pandemic has caused significant upheaval in the lives of U.S. citizens. The pandemic has also become highly politicized, with each political party criticizing the other for its response (or lack thereof) to the pandemic, including stay at home orders, closures of local businesses, and the encouragement of masks and social distancing. Simultaneously, U.S. citizens have been forming their own opinions about local and federal responses, and in this report we analyze if citizens' opinions of their elected governors differ depending on whether or not someone in their household has tested positive for COVID-19. Knowing this information prior to the 2020 election can help state governments and political parties understand the opinions of their constituents, which could influence their campaign strategies. To test if the approval rate of governors is different between households that have tested positive for COVID-19 versus those that have not, we utilize comprehensive data from the American National Election Studies (ANES) 2020 Time Series Study and a Wilcoxon rank sum test.

2 Description of Data

The ANES data set contains information from 8,280 pre-election interviews with U.S. citizens of voting age. Two variables are particularly relevant for us to answer this question:

- `V201145`: APPROVE OR DISAPPROVE R'S GOVERNOR HANDLING COVID-19
- `V201624`: ANYONE IN HOUSEHOLD TESTED POS FOR COVID-19

Both variables are binary variables. In `V201145`, governor approval is represented by the value 1, while disapproval is represented by the number 2. We remove the remaining irrelevant values, including those for refusal to answer and interviewees whose responses got cut off. The breakoff of a survey response could happen for a variety of reasons, including lack of Internet or phone service.

For `V201624`, the value 1 is assigned if there has been a positive COVID-19 test in the respondent's household; the value 2 is assigned if there has been no positive COVID-19 test in the respondent's household. We remove the irrelevant values corresponding to those who refused to answer.

After cleaning the data, we are left with 8,138 valid observations. We can see from the data summary below that our variables are in the expected ranges.

```
summary(df_clean)
```

```
##       gov            covid
##  Min.   :1.000   Min.   :1.000
##  1st Qu.:1.000   1st Qu.:2.000
##  Median :1.000   Median :2.000
##  Mean   :1.381   Mean   :1.965
##  3rd Qu.:2.000   3rd Qu.:2.000
##  Max.   :2.000   Max.   :2.000
```

We additionally do some initial data exploration. Table 1 shows how our data is divided among our two Boolean variables. It is clear that the vast majority of households have not had a positive COVID-19 test. However, further analysis about how this relates to governor approval is less clear. Figure 1 helps visualize how the data is split. The pie chart on the left shows the approval/disapproval rates for the households that have received a positive COVID-19 test, and the pie chart on the left is for households that have not had a positive test. We can see that there is some difference in how the approval and disapproval rates are split– on average, 44% of households that have tested positive disapprove of their governor's response, while this value is only 38% for households with no positive COVID-19 test. We want to test if there is a statistical significance in this difference, and our methods are described further in the next section.

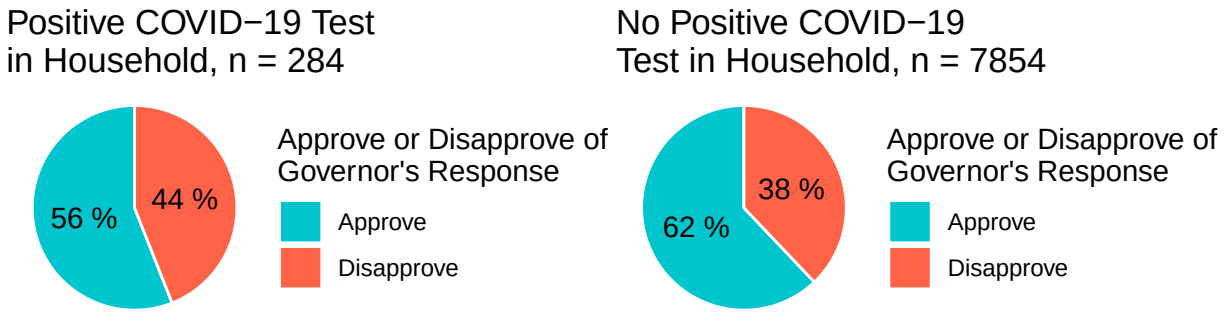

Figure 1: Distribution of Voters' Approval of Governor based on Positive COVID-19 Test in Household

Table 1: Cross Tab of Governor Approval and Positive COVID-19 Tests

	No Positive Test	Positive Test
Approve	0.599	0.020
Disapprove	0.366	0.015

3 Most appropriate test

First, we notice that the governor approval variable is a Boolean variable. There are only two valid answers: Approve, or Disapprove. As a result, a parametric test based on some underlying distribution that resembles Normal would not be appropriate.

At the same time, the groups (people who had a positive test in their household v.s. people who did not) are distinct people, and they don't have a natural pairing.

Furthermore, given the sampling frame based on a cross-section of registered addresses across 50 states and the District of Columbia, we feel the data are sufficiently close to be i.i.d.

Based on the above diagnose, the Wilcoxon Rank Sum test is the most appropriate in this case.

4 Test, results and interpretation

We establish the *null hypothesis* to be that the average support for the respondent's governor is the same among people that had a member of the household test positive for COVID-19 and those who did not. Given we have no strong initial inclination in either direction, this should be a two tailed test. We use the standard 5% significance level.

First take a look at the table in each of the 4 cases for context, then run the test.

```
table(df_clean$approval, df_clean$covid_tested)
```

```
##
##              Not positive Positive
##   Approve            4876      159
##   Disapprove         2978      125
```

```
wilcox.test(df_clean$gov ~ df_clean$covid)
```

```
##
```

```
##  Wilcoxon rank sum test with continuity correction
##
## data:  df_clean$gov by df_clean$covid
## W = 1183267, p-value = 0.0377
## alternative hypothesis: true location shift is not equal to 0
```

From the test we can see that the p-value is 0.0377, which is less than the significance level $\alpha = 0.05$, meaning that we reject the null hypothesis in favor of the alternative that people with a positive COVID-19 test in their household have a different opinion of their Governor than people without.

Practically, we can calculate the correlation between the two variables.

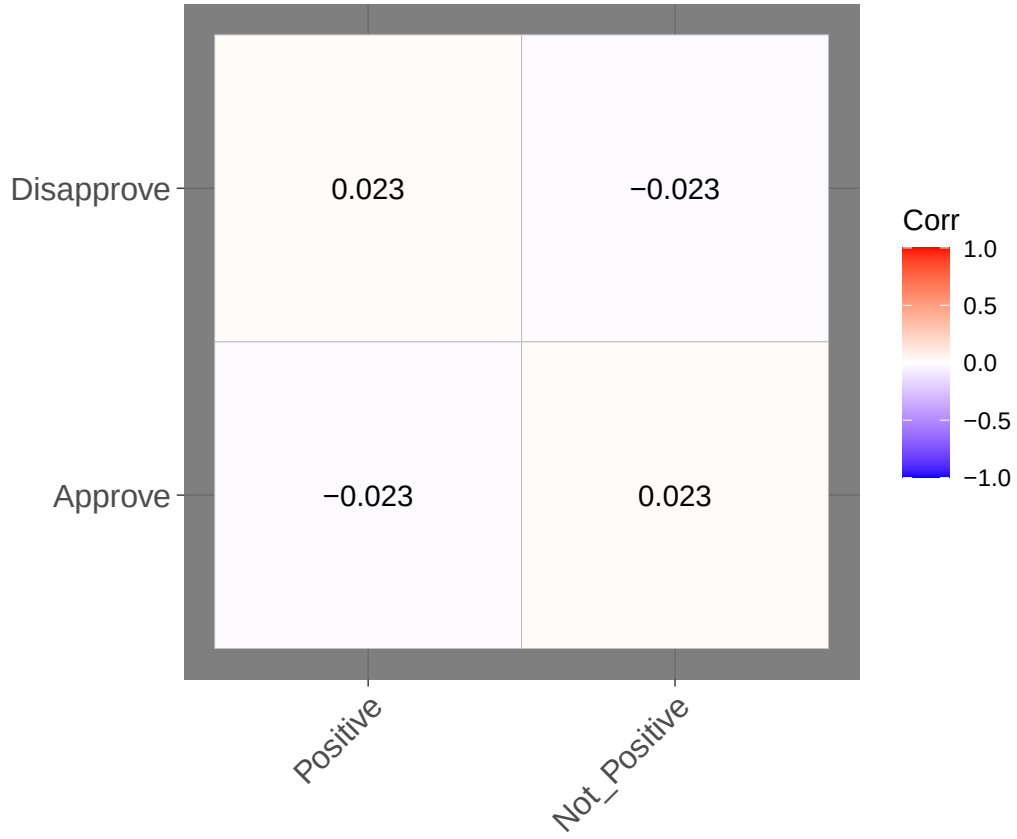

Figure 2: Correlation Table for Voters' Approval of Governor and Household COVID-19 Tests

As presented in Fig. 2, having a positive test in the household does linearly correlate to less likely to approve the Governor's handling of the pandemic. However given the magnitude of the the correlation coefficient. The linear relationship isn't very strong.