# 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. 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. This report analyzes if citizens' opinions of their elected governors differ depending on whether someone in their household has tested positive for COVID-19. Understanding this pre-2020 election information election can help state governments and political parties identify their constituents' opinions, turning their future campaign strategies in a different direction. To test if governors' approval rate is different between households that have tested positive for COVID-19 versus those who have not, we utilize comprehensive data from the American National Election Studies (ANES) 2020 Time Series Study 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

While there are other variables in the ANES data set that could have been helpful in answering this question, including `V201146 -- PRE: HOW MUCH APPROVE/DISAPPROVE R'S GOVERNOR HANDLING COVID-19`, we believe the two highlighted above best address and most successfully operationalize our research question.

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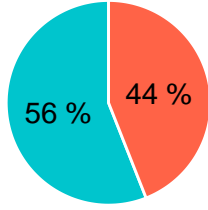
After cleaning the data, we are left with 8,138 valid observations. We can see from the data summary below that our variables are in the expected ranges.

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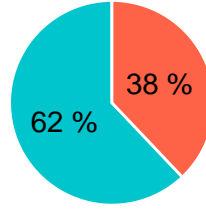
Figure 1: Distribution of Voters' Approval of Governor based on Positive COVID-19 Test in Household

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

To answer our question of whether individuals' opinions of their governor's response to the COVID-19 pandemic depends on whether someone in their household has tested positive for COVID-19, we use a Wilcoxon Rank Sum Test. We believe this test is most appropriate for the given data because our data is ordinal: approval/disapproval of the governor's response is a categorical variable with a natural ordering. A parametric test based on some underlying distribution that resembles a normal distribution would not be appropriate. Furthermore, the two groups of interest (people who had a positive test in their household v.s. people who did not) are distinct observations, and they don't have a natural pairing.

There are two main assumptions for the Wilcoxon Rank Sum Test:

1. Ordinal variable: As discusses already, approval/disapproval of the governor's response is an ordinal variable.

2. i.i.d.: 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.

With all assumptions sufficiently met, we proceed with a Wilcoxon Rank Sum Test.

# 4 Test, results and interpretation

We establish the *null hypothesis* to be that there is no difference in the opinions of the governor's response between individuals who have had a positive COVID-19 test in their household compared to those who have not had a positive test. Our *alternative hypothesis* is that there is a difference between the opinions of the two groups. Given we have no strong initial inclination in either direction, we conduct a two tailed test. We use the standard 5% significance level.

```
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 $\alpha = 0.05$.We reject the null hypothesis in favor of the alternative that people with a positive COVID-19 test in their household do have a different opinion of their governor's response than people without.

There is little practical significance in these results. Because we have a large sample size, we calculated the Wilcoxon Rank Sum effect size. We found the effect size to be equal to 0.023, which is a very low value and indicates a very low-magnitude effect. Practically speaking the difference between the two groups in our test is very minimal, and our pie charts in Figure 1 further corroborate this finding. Interestingly, however, it is likely that politicians and political parties made specific appeals in their campaigns to families who have received a positive COVID-19 test in hopes of winning their vote due to their dissatisfaction with their governor's or state's response to COVID-19. Even this small difference likely caused a change in campaign strategies.