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?

Elaine Chang Dom Dillingham Jesse Miller Michael Wang

1. Introduction

2020 was marked by the novel COVID-19 pandemic. Nearly all Americans experienced changes to their lives, but perhaps those whose lives were most affected were those with a COVID-19 infection themselves or within their immediate households. In early November 2020, the US had recorded 10 million confirmed COVID-19 cases. How do these infected Americans and their households feel about how their elected officials handled the pandemic?

In this analysis, we test whether survey respondents who had someone in their household infected by COVID-19 are more likely to disapprove of the way their governor is handling the pandemic when compared with respondents without household infection of COVID-19. This research question poses far-reaching implications on the democratic process. 2020 was a crucial election year, during which gubernatorial elections were held in 11 states and 2 territories, concurrently with the presidential election. A stark change in perception of a governor's ability to navigate a crisis could swing leadership in up to 9 states where governors ran for reelection, and affect the political balance in more states during mid-term elections.

2. Data

Raw data originates from the 2020 Time Series Study, a pre-election survey that the American National Election Studies (ANES) conducted on 8,280 eligible voters in the United States. Each survey question corresponds to a variable in the dataset. Within this dataset of more than 700 variables, 4 were identified to be relevant to this analysis:

1. V201624: Has anyone in your household tested positive for the coronavirus disease, COVID-19, or has no one tested positive?
2. V201625: Has anyone in your household been suspected of having COVID-19 based on their symptoms, or not?
3. V201145: Do you approve or disapprove of the way [governor] has handled the COVID-19 pandemic?
4. V201146: Do you [approve/disapprove] strongly or not strongly?

We remove entries in which survey respondents failed to answer the above questions with a binary response. These instances included when the survey respondent broke off an interview, refused to answer, or responded with "Don't know" or "Inapplicable". These removed entries constituted 1.7% of the raw data. We believe the nature of this non-response is random, not systematic, so we believe the removal of these entries will not bias the results of our analysis. We also normalize binary responses onto a scale of 0 to 1.

We operationalize a respondent who has someone in their home infected by COVID-19 by examining two variables: V201624 and V201625. Although similar, the former variable depends on the result of a COVID-19 test, while the latter variable simply asks the respondent of their suspicion of a household member with COVID-19 due to symptoms. In this analysis, we choose to analyze respondents under the lens of the former, test-driven variable, V201624. Although we recognize this may bias our results against respondents who are unable to receive a COVID-19 test, we believe relying on the ground truth - a positive or negative test result - is most appropriate.

We operationalize a respondent's perception of how their governor handles COVID-19 using two variables: V201145 and V201146. The former variable records a binary response (approval or disapproval), while the latter variable records the magnitude of response (strongly or not strongly). We create an additional ordinal variable that combines a respondent's binary approval of a governor with the respondent's strength in opinion. This variable serves as a Likert scale variable with 4 categories: strongly disapprove, disapprove, approve, and strongly approve.

No weighting was applied in this analysis. There are flaws to this assumption, which will be discussed in *Section 4: Significance.*

Basic observations of the data indicate 284 respondents with someone in their household who tested positive for COVID-19, of which 44% disapproved of their governor's handling of the pandemic. Conversely, of the 7,854 respondents without a positive COVID-19 case in their household, 38% disapproved of their governor's handling of the pandemic. From a cursory analysis, the means of the data indicate survey respondents who have had someone in their home infected by COVID-19 may be more likely to disapprove of the way their governor is handling the pandemic, but statistical tests need to be performed in order to evaluate statistical significance.

Table 1: Households With a Confirmed COVID Case

COVID Case in Household?	Approve of Governor?	Count	Percent
Yes	Approve	159	56
Yes	Disapprove	125	44

Table 2: Households Without a Confirmed COVID Case

COVID Case in Household?	Approve of Governor?	Count	Percent
No	Approve	4876	62
No	Disapprove	2978	38

3. Tests

We approach the research question through the interpretation that disapproval exists on a spectrum. For example, two respondents may report they "Approve" and "Strongly Approve" of their respective governors. While both respondents approve of their governors, the former respondent disapproves of their governor to a relatively greater degree than the latter respondent who approves more strongly. We aim to capture this ordinal relationship by implementing a non-parametric Wilcoxon Ranked Sum test.

A non-parametric Wilcoxon Ranked Sum (WRS) hypothesis of comparisons test accounts for respondents' relative approval of how their governor handles COVID-19. Our sample data is split into two groups: those with household members who tested positive for COVID-19, and those without. Relative approval of the respondent's governor is defined by a custom categorical feature that contains the values: strongly disapprove, disapprove, approve, and strongly approve. This feature is generated by combining the binary approval or disapproval rating respondents report (V201145) with their strength in opinion (V201146).

Table 3: Likert Approval of Governor's Handling of COVID-19

	COVID-positive	COVID-negative
Strongly Approve	95 (33%)	3087 (39%)
Approve	64 (23%)	1780 (23%)
Disapprove	40 (14%)	843 (11%)
Strongly Disapprove	85 (30%)	2134 (27%)

A WRS hypothesis of comparisons tests is appropriate because it meets the following core assumptions:

1. **Ordinal data.** The custom variable we create, which combines the binary nature of V201145 and the magnitudes of V201145, is an ordinal data type. We cannot, and do not attempt to, make comparisons beyond the $<$, $>$, or $=$ operators, because we recognize the intervals between categories are subjective.
2. **Independently and identically distributed (IID) data draws.** This assumption is likely violated to some degree. It is highly unlikely all survey respondents are truly independently selected. The ANES Guidebook states a large subset of survey respondents are return respondents from the 2016 ANES survey, which indicates survey respondents in 2020 have a dependence upon prior affiliation with ANES.

Additionally, it is unlikely survey respondents are identically distributed across the population of all eligible US voters, especially as no weight adjustments have been applied in this analysis. Despite these violations, we have a large sample of ~8,000 survey respondents who have been sampled across 50 states, and we operate under the assumption that our data does not violate the IID assumption too badly. Other hypothesis tests also require an IID assumption, so we believe a WRS test violates this assumption no more than other tests.

Although we notice the size of the with-COVID (284) and without-COVID (7,854) groups differ by a substantial margin, the Ranked Sum Test will normalize for the difference in sample sizes across the two groups. We opt for a WRS test over a dependent Wilcoxon Sign Rank Test or Sign Test because there exists no logical pairing across the two groups. We opt for a WRS test over a z-test because we have minimal knowledge about the population distributions of the two groups. We opt for a WRS test over a t-test because we aim to capture the respondents' strength in opinion of their governor, and a t-test would fail to account for this ordinal relationship.

We operationalize our test according to the following hypotheses:

- Ho: The probability of a draw from the with-COVID group ranking higher than a draw from the without-COVID group is equal to the probability of a draw from the without-COVID group ranking higher than a draw from the with-COVID group.
- Ha: The probability of a draw from the with-COVID group ranking higher than a draw from the without-COVID group is not equal to the probability of a draw from the without-COVID group ranking higher than a draw from the with-COVID group.

Assuming an industry-standard significance level, α, of 5% across two tails, a WRS test yields a p-value of 0.042. Because this p-value is less than our significance level, the test implies a statistically significant finding that the null hypothesis should be rejected. Specifically, the probability of a with-COVID respondent holding a more disapproving opinion of their governor is greater than the probability of a without-COVID respondent reporting a similar degree of disapproval. The with-COVID group has a mean disapproval of 44%, while the without-COVID group has a mean disapproval of 38%. We calculate a Cohen's d estimate, a measure of the difference between sample means in standard deviation units, of -0.12. This implies negligible significance. Similarly, we compute a correlation of -0.022, which indicates a weak correlation between the two groups.

4. Significance

Although the Wilcoxon Ranked Sum test reports a statistically significant difference between with-COVID and without-COVID groups, we believe the practical implications of this finding is questionable for two reasons.

Firstly, although a 6-point margin can result in non-negligible changes to a governor's public image, this magnitude is small when compared to typical fluctuations in approval ratings that elected officials experience. A comparison can be made to the changes in approval ratings of the Office of the President, which is well documented. Specifically, we examine the largest magnitude change in approval rating over a president's term. Over the past 5 US presidents, there was an average approval rating spread of 41%. Stated differently, among the last 5 US presidents, each has seen an average 41% fluctuation in their approval rating during their term. If we use fluctuations in presidential approval ratings as an analog for fluctuations in governor approval ratings, a 6% change in the midst of an anomalous COVID-19 pandemic would likely carry non-lasting impacts to a governor's political future. We find that in either case (with-COVID or without-COVID) the majority of the two populations approve of their governor's handling of COVID-19.

Secondly, our sampling method is highly vulnerable to bias. Bias occurs when the expected value of our estimator is not equal to the true value in the population ($E[\hat{\theta}] \neq \theta$). This commonly occurs when one is sampling from a group that does not truly represent the population. This risk exists in our analysis. The ANES dataset takes steps to ensure its respondents represent a random sample of the population of all eligible American voters, but we have not applied the weights and adjustments that ANES recommends. Furthermore, when we divide the ANES dataset into with-COVID and without-COVID groups, our samples may no longer reflect true random draws from the population. To illustrate this problem further, our sample of 284 respondents who had a confirmed COVID-19 case in their household could have been overdrawn from a small handful of states with a particularly disliked governor, and would not reflect on how the overall population of Americans who had a confirmed COVID-19 case in their household would perceive their leadership's effectiveness at handling COVID-19.

5. Appendix

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  test_positive and test_negative
## W = 1038950, p-value = 0.04289
## alternative hypothesis: true location shift is not equal to 0
```

```
##
## Cohen's d
##
## d estimate: -0.1190505 (negligible)
## 95 percent confidence interval:
##         lower         upper
## -0.2374714224 -0.0006295166
```