

# W203 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?

Tilman Bayer, Kieran Berton, Ruilan Liu, Toby Petty

## Importance and Context

The COVID-19 Pandemic, first identified in Wuhan, China in December 2019, has caused over 500,000 deaths as of March 2021 in the United States, the highest death count of any nation in the world. Though the federal government labelled the outbreak a public health emergency on January 31st, 2020, responses by state governments have varied, with some imposing stringent lockdowns and curfews, others closing business and public spaces, and some imposing facemask mandates in public.

Despite this, the federal and state governments have faced sharp criticism for their handling of the pandemic, as infection rates and death tolls worsened for many states following the lifting of Coronavirus-related restrictions. Nevertheless some have defended the actions of state and local governments, citing the need for businesses to be open to serve public needs. In this study, we hope to learn more about the connection between a person's level of exposure to the virus and their opinions on how state governments have handled the pandemic. We suspect that a person who contracts or lives with someone who contracts the Coronavirus is more likely to assign some of the blame for their circumstance to the government's neglect of public health concerns.

Using data from the American National Election Study (ANES) 2020 Pre-Election Survey, we will compare approval for how governors have handled the pandemic between a population of respondents who have not been exposed to the virus and a population of respondents that either have contracted the virus or live with someone who has contracted the virus. As governors' responses have varied on a state by state basis and we are looking at respondents on the national level, we cannot cite any particular actions that may have contributed to our results, nor can we make statements about the connection between public approval and state coronavirus responses for any particular state. However, the answer to this question could provide insight into the connection between the US population's opinions on coronavirus restrictions and citizens' level of exposure to the virus. We cannot make any statements about the causal connection between these two variables, as this data was merely observational and was collected from respondents at one moment in time, but we hope to gain a better understanding of how the two variables are correlated through this experimental setup.

## Description of Data

Please refer to our report for question 1 for general remarks about the ANES dataset,<sup>1</sup> As stated there, the ANES dataset contains responses from 8280 respondents to approximately 700 questions concerning their political opinions, voting behaviors, as well as some qualitative questions about their opinions on social issues.

We find 4 relevant variables in the dataset that can answer the question:

*V201145 - PRE: Approve or disapprove R's governor handling COVID-19*

*V201146 - PRE: How much approve/disapprove R's governor handling COVID-19*

---

<sup>1</sup>To call out one aspect here: As advised by ANES, an accurate analysis based on this data would have to account for the survey weights included in the dataset. Per the instructors, we are ignoring them for the purposes of this labs project.

Table 1: Summary of Responses

V201145 code	V201145 responses	V201624 code	V201624 responses
-9	36	-9	60
-8	16	-5	38
1	5083	1	287
2	3145	2	7895
		V201625 code	V201625 responses
		-9	59
		-5	38
		1	998
		2	7185

*V201624 - PRE: Anyone in household tested pos for COVID-19*

*V201625 - PRE: Anyone in household COVID-19 based on symptoms*

### Shape data

Because the question is particularly interested in the likelihood of disapproval, not the degree of disapproval, we pick V201145 over V201146 in our analysis. In V201145, there are 2 codes, -8 (don't know) and -9 (refused), that are not informative about our analysis, so they are dropped. These data only count for a very small portion(5.46%) of the sample, as shown in Table 1, and won't largely change the result. The same holds true for V201624 and V201625 that -5(interview breakout) and -9(refused) could be excluded in our analysis. After reshaping the data, there are 8132 responses remaining.

### Re-encode data

Now that we have V201145 responses in only code 1(approve) and 2(disapprove), we can re-encode it into 0 (approve) and 1 (disapprove) to become a dummy variable so that we could convert them into metric numbers and do a t-test directly without manipulations.

### COVID vs. non-COVID group

In the next step, we need to identify the group that was infected with COVID-19 so as to compare with the non-infected group. We decide to put respondents marked as 1 (Someone tested positive) in V201624 and 1 (Suspected of having COVID) in V201625 in the COVID group, because people who suspected of having COVID based on symptoms also went through a hard time even without testing positive and as a result, they could also hold negative opinions on their governors' handling of pandemic. Besides, due to the limited public resources, many people suspected to have COVID didn't get a chance to do the test. So we figure it'd be appropriate to include both variables into our analysis. As a result, the COVID group consists 1065 respondents while the non-COVID group has 7067 people.

### Most appropriate test

We are going to use a two-tailed t-test, with  $\alpha = 0.05$  (a standard value used in many political science analyses).

- **Null Hypothesis:** the ratios of disapproval among the COVID group and the non-COVID groups are the same. This ratio is equal to the mean of the variable we constructed above.
- **Alternative Hypothesis:** the difference between those ratios is non-zero.

The usual assumptions for such a t-test are satisfied:

- Since our (formerly categorical) random variable only takes on two values (0 and 1), we can regard it as a metric variable.

- We can assume that the responses are i.i.d (with some minor limitations, e.g. respondents from the same household would obviously not have independent answers).
- While the distribution is not normal (concentrated on 0 and 1), our sample size is sufficiently large to justify using a t-test.
- We use a two-tailed test because we cannot rule out that people who had someone in their home infected could be *less* likely to disapprove of their governors' way of handling the pandemic.

We think this is the most appropriate test to use in this situation. A non-parametric test would have the disadvantage of lower statistical power (i.e. being less likely to reject the null hypothesis if it is not true). And since there is no natural correspondence between individuals in households with and without infections (and we don't have data allowing us to compare the same individual's opinion before and after an infection occurred in their household), a paired test would not make sense.

## Test, results and interpretation

```
test <- t.test(covid$V201145, non_covid$V201145, alternative = "two.sided")
```

The two-tailed t-test result shows that the mean or the disapproval rate of the COVID group is 0.4 while the mean for the non-COVID group is 0.38. 0.02 represents a difference where p-value = 0.18 is well beyond alpha=0.05. In other words, we fail to reject the null hypothesis that the difference of the two means are 0.

Based on our test, we would say that the survey respondents who have had someone in their home infected by COVID-19 show no statistically significant difference in disapproval of the way their governor is handling the pandemic compared with the ones who don't.