

Lab 1: Question 3

Amanjot Samra, Leena Bhai, Jessica Sacks

Contents

0.1	Importance and Context	2
0.2	Description of Data	2
0.3	Most appropriate test	2
0.4	Test, results and interpretation	3

0.1 Importance and Context

The COVID-19 pandemic has permanently altered the world as we know it. With over 25 million cases and hundreds of thousands of deaths in the US alone, the government has come under criticism for its management, or lack thereof, to slow the spread of covid. Governors have been at the forefront of the US's battle against this pandemic, and thus, each state has handled its covid outbreaks differently. As of 3/2/2021, the governor of Texas has fully lifted all covid-related restrictions while some states remain in a strict lockdown and some never went into lockdown in the first place.

The life of every single individual in the United States has been affected in some way, regardless of whether or not they actually contracted the virus. Consequently, it is interesting and worth investigating the overall sentiment toward the handling of the pandemic response at the state government level, and specifically investigate if the presence of covid infections in one's family are a contributing factor to them disapproving of their governor.

0.2 Description of Data

We will answer our research question using the 2020 American National Election Studies data. The data in this database was collected before the 2020 elections. Respondents were drawn from the general US population at random, with the requirement being that they are eligible and registered to vote. The questions asked to respondents that are relevant to this question are:

1. Do you approve or disapprove of the way [Governor of respondent's state] has handled the COVID-19 pandemic?
2. Has anyone in your household tested positive for COVID-19 or has no one tested positive?
3. Has anyone in your household been suspected of having COVID-19 based on their symptoms, or not?

Respondents had the option of replying yes/no and some form of "refused/inapplicable/no answer". We filtered out the later of these and focused only on those who said yes or no in order to remove incomplete data points from the data set. Our question is to find out if survey respondents who have a household member are more likely to disapprove of the way their governor is handling the pandemic, and so we plan to break our data into to groups with the grouping variable being covid infections in the household: those with no positive covid tests and no covid symptoms in their household for the first group, and those with either positive covid tests or covid symptoms in their household in the second group. We combined covid positive tests and covid symptoms into one binary variable for the purposes of this research.

##			
##		Approves of Governor Disapproves of Governor	
##	No covid in household	4398	2680
##	Covid infection in household	641	426

The above table displays the counts of people who fall into each of the four categories: covid-infected house + approves of governor, covid-infected house + disapproves of governor, covid-free house + approves of governor, covid-free house + disapproves of governor. The 2680/7078 or 37.86% respondents in covid-free households disapprove of their governor, while 426/1067 or 39.92 % of respondents with a covid-infected household member disapprove of their governors.

Note: One limitation of this data is that there are 50 governors and the data we have received is not broken down by governor. Therefore, it is possible that we have a higher amount of respondents from one state versus another, and that could possibly skew our analysis, for example if there is one state with a higher number of participants with a particularly high disapproval or approval rating of their governor.

0.3 Most appropriate test

Our outcome variable for this question is a binary variable, disapproval of the governors. It's binary because the only two choices are that the respondent approves of how the governor is handling the COVID-19 pandemic or the respondent disapproves. We can treat this binary variable as a metric variable if we recode

the responses as a 0 or 1, and then our interval will be from 0 to 1. We will combine people who are covid positive or have covid symptoms into one variable, a 0 if neither are true and a 1 if either are true. We will remove from our dataset any respondents who refused to answer, or answered with inapplicable or don't know for the question of governor approval. After this, we can evaluate metrically the disapproval/approvals within each group (covid/no-covid) with a two sample t-test, with the objective of the test being to determine whether the difference between these two groups is statistically significant.

Assumptions for a two sample t-test:

1. metric scale: a binary variable is metric because there is only one interval
2. iid data: Although the population of respondents may not directly reflect the entire population of the United States, the data is identically and independently distributed. We can examine the sample's iid in the sample design and respondent recruitment for the survey. The survey was conducted three ways: self-administered online survey, live video interviews, and telephone interviews. It is identically distributed in that each respondent is drawn from the population of interest, the US citizens aged 18+. Independence is met in that sampling occurred from a random draw from the USPS delivery addresses across the US. The mode of survey was also computer randomized.
3. The distribution is bimodal and we have a large sample size.

We will use a p-value of .05.

0.4 Test, results and interpretation

Null hypothesis: Covid infected households and non-covid infected households are equally likely to disapprove of their governor's handling of the pandemic.

Alternate hypothesis: There is a relationship between covid-infections in one's household and their likelihood to disapprove of their governor's handling of the pandemic.

```
t.test(Q3$DisapprovesGovernor ~ Q3$HasCovid, alternative = "two.sided")
```

```
##
## Welch Two Sample t-test
##
## data: Q3$DisapprovesGovernor by Q3$HasCovid
## t = -1.2827, df = 1399.7, p-value = 0.1998
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.05213603 0.01091163
## sample estimates:
## mean in group FALSE mean in group TRUE
## 0.3786380 0.3992502
```

Since the p-value of the t-test we conducted is greater than .05, we fail to reject the null hypothesis that covid infected and covid-free households are equally likely to disapprove of their governor's handling of the pandemic. Comparing the means, we can see that the fraction of people with covid-infected household members who disapprove of their governor's handling of the pandemic is $0.3992502 - 0.3786380 = 0.0206122$ higher than the fraction of people without covid-infected household members who disapprove of their governor's handling of the pandemic.

Thus, the outcome of this test is not statistically significant. Furthermore, the outcome of this test is also not practically significant. There are many factors other than covid-infected household members, or the absence thereof, that are contributing to the approval or disapproval of governors. For example, people whose businesses or means of income are being negatively impacted by the pandemic may have very negative views of their governor regardless of whether or not someone in their household has contracted the virus. Healthcare workers will have strong opinions about this issue regardless of their own household as well.