

Digging Deeper into COVID Impact Survey

Introduction

For our group project, we looked over a COVID data set from canvas. The two main questions we asked were: does living with someone who has a positive case of COVID increase the odds of you contracting the disease? Does the number of people living in your house increase the chances of contracting COVID

We used variable PHYS4 and ran a frequency table to determine how many people in the study contracted COVID. Then we used two other variables PHYS5 and HHSIZE1 to see if and how those variables impacted the PHYS4 variable.

By seeing how the PHYS5 and HHSIZE1 variables impacted the PHYS4 variable, we were able to draw conclusions about the sample and how these different variables impacted the chances of the respondent getting COVID

Methods

- We used the freq function to generate frequency tables
- Barplot function to create our univariate and bivariate graphs
- Chi squared test of independence to conclude that there was a relationship between the variables PHYS4 and PHYS5

Results

- We used the Chi squared test of independence to compare two of our categorical variables. The variables were whether or not the respondent lived with someone that was diagnosed with COVID (explanatory) and whether or not the respondent themselves was diagnosed with COVID. We expected that these would be correlated since a respondent is more likely to contract a virus if someone they live with did, and conducting the test was a way to prove this.
- The syntax that was run was:
- `ND.chisq<-chisq.test(data$PHYS4,data$PHYS5)`
- `ND.chisq`
- The results gave us an χ^2 value of 10478, and a p value of 2.2e-16.
- Since the p value is less than .0001, we can conclude that the variables are correlated. Since the explanatory variable is only 2 values, a post hoc test is not necessary.

The screenshot shows the RStudio interface with several panes:

- Code pane:** Displays R code for data cleaning and frequency analysis.
- Console pane:** Shows the execution of the R code, including the summary of the dataset and the frequency distribution of the variable `data$phys5`.
- Environment pane:** Shows the loaded datasets: `covid_w3_delivered` and `data`, both with 7505 observations and 177 variables.
- Plots pane:** Displays a bar chart of the frequency distribution of `data$phys5`. The x-axis categories are "(1) Yes", "(2) No", "(77) Not sure", and "(99) REFUSED". The y-axis represents Frequency, ranging from 0 to 7500. The bar for "(1) Yes" reaches approximately 7200, while others are near zero.

The figure shows the RStudio interface with several panes:

- Code pane:** Displays R code for data cleaning and frequency distribution calculation.
- Environment pane:** Shows the global environment with two datasets: COVID.WL.delivered and data.
- Console pane:** Displays the output of the R code, including frequency counts for household size.
- Plots pane:** Displays a bar chart of household sizes.

Code pane content:

```
library(magick)
library(desctools)
library(ggplot2)
library(tidyverse)
library(dplyr)
library(rstatix)
data<-COVID_WL_delivered

#freq distribution of people who were diagnosed with covid
#first we are going to code out the missing data (refused, skipped)
data$PERS4 [data$PERS4==77]<-NA
data$PERS4 [data$PERS4==98]<-NA
data$PERS4 [data$PERS4==99]<-NA
freq(data$PERS4)

data$HHSIZE1 [data$HHSIZE1==77]<-NA
data$HHSIZE1 [data$HHSIZE1==98]<-NA
data$HHSIZE1 [data$HHSIZE1==99]<-NA
freq(data$HHSIZE1)
```

Console output:

```
(77) ONE PERSON          43  0.07233
(98) SKIPPED ON WEB      15  0.19887
(99) REFUSED              1  0.00132
Total                  7505 100.00000
> data$HHSIZE1[data$HHSIZE1==77]<-NA
> data$HHSIZE1[data$HHSIZE1==98]<-NA
> data$HHSIZE1[data$HHSIZE1==99]<-NA
> freq(data$HHSIZE1)
data$HHSIZE1
```

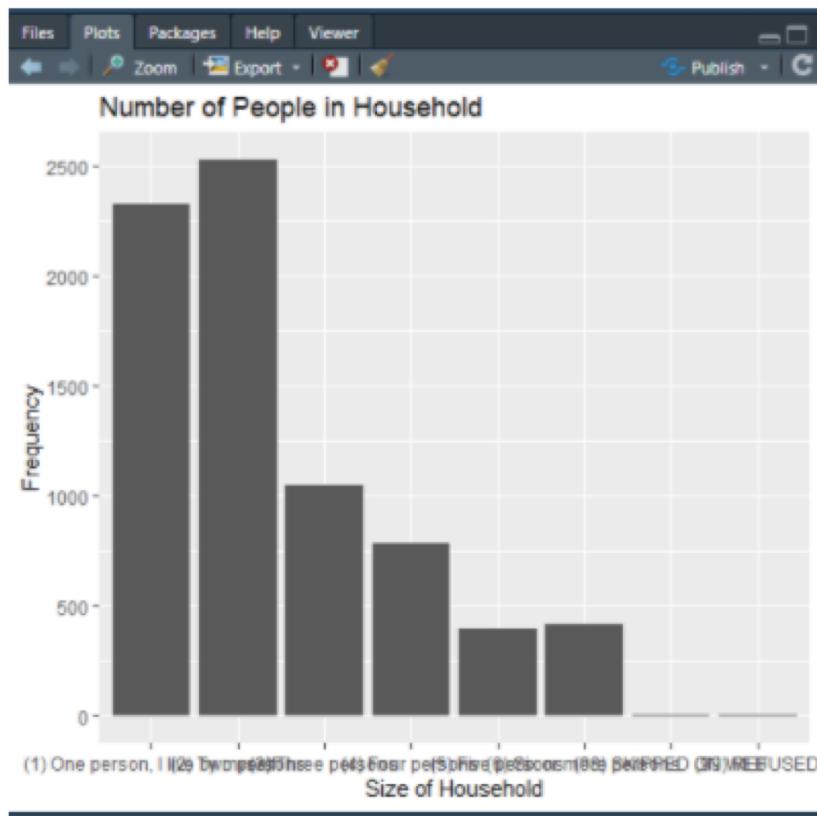
	Frequency	Percent
(1) one person, I live by myself	2324	30.9660
(2) two persons	2326	31.6378
(3) three persons	1050	13.9907
(4) four persons	781	10.4064
(5) Five persons	398	5.2763
(6) six or more persons	418	5.5430
(98) SKIPPED ON WEB	8	0.1064
(99) REFUSED	4	0.0531
Total	7505	100.0000

Plots pane content:

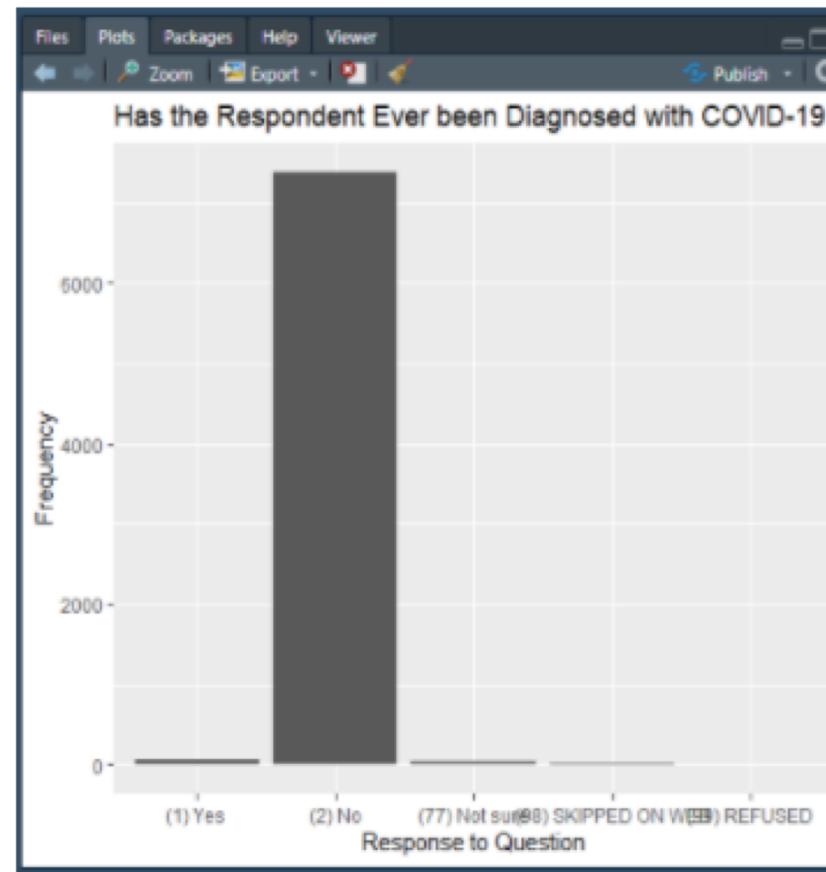
A bar chart showing the frequency distribution of household sizes. The x-axis categories are labeled as follows:

- (1) One person, I live by myself
- (5) Five persons
- (99) REFUSED

The y-axis represents frequency, ranging from 0 to 2500. The bars show frequencies of approximately 2300, 2500, 1000, 800, 500, and 0 for the respective categories.

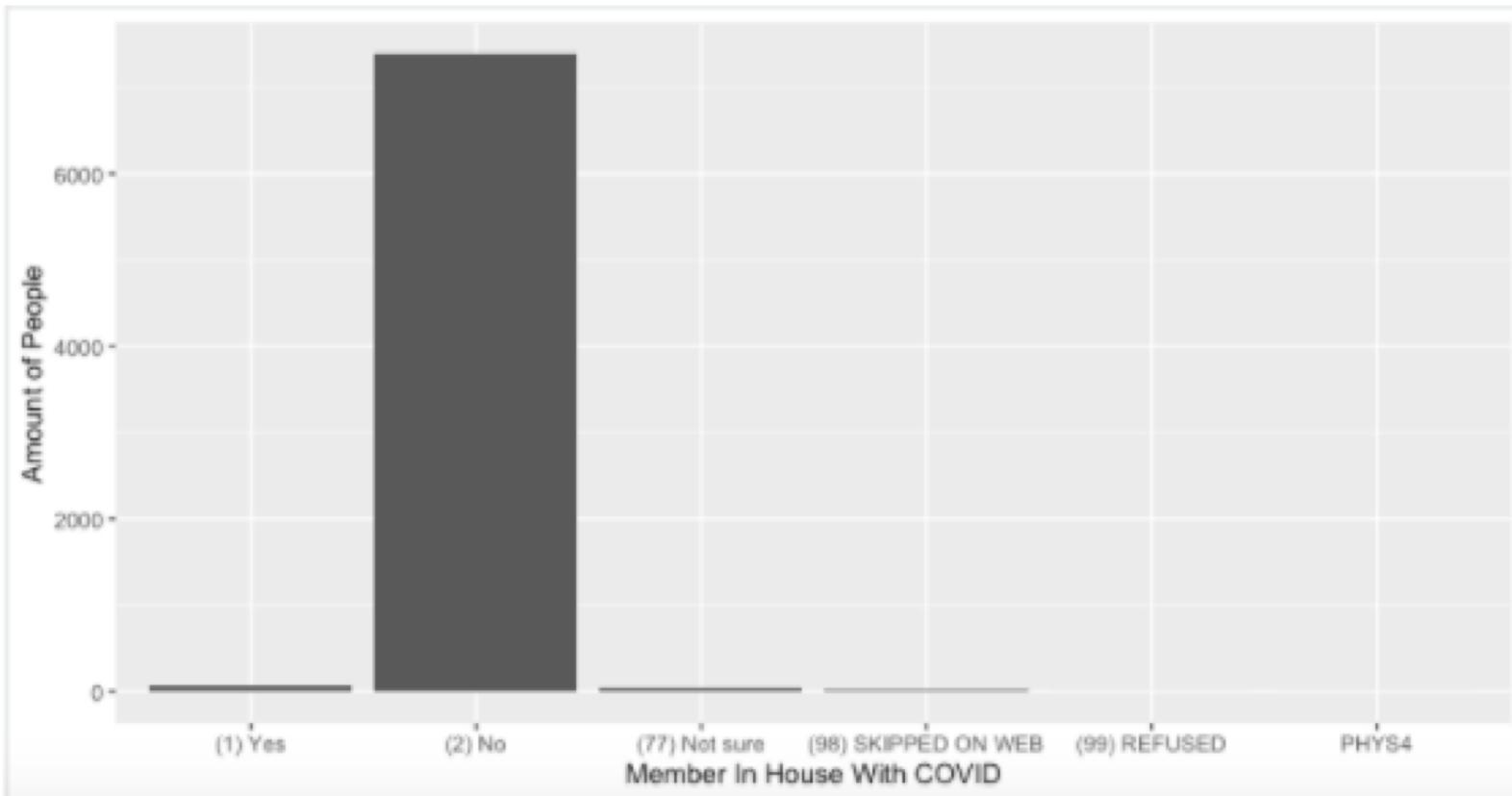


This bar graph represents the number of people per household that participated in the survey (HHSIZE1). The graph is clearly skewed to the left which shows that it is much more likely for members of small households to be taking part in the survey, rather than households with a lot of residents. The most frequent category is 2, which represents two people living together in a house. The least popular category is five persons.



The second bar graph here represents the variable PHYS4, which is whether or not the respondent has ever been diagnosed with COVID-19. Clearly, the results are very much skewed towards the response no. This tells us that our sample was likely not one that experienced an outbreak or was at high risk of contracting the virus.

This next bar graph represents the variable PHYS5. The results show the amount of people living with some member of their house who was diagnosed with COVID. The results are clearly skewed towards "NO", so the majority of people were not living with someone who was diagnosed with COVID.



Conclusion

- In conclusion there is correlation between the PHYS4 variable and the PHYS5 variable. The bar graphs for both variables were skewed far right, and the p-value showed us that there is a correlation between whether or not the respondent lived with someone who was diagnosed with COVID and whether or not they received it themselves.

Thanks For Watching!

- Randy Cory- u1158094@Utah.edu
- Gabby Sosnowski- gabbysosnowski18@gmail.com
- Jackson Schor- u1115937@utah.edu

Citations

- COVID_data.csv set from Canvas
- Bowleg, L. (2020). We're Not All in This Together: On COVID-19, Intersectionality, and Structural Inequality. *American Journal of Public Health*, 110(7), 917-917. doi:10.2105/ajph.2020.305766
- Dunn, C. G., Kenney, E., Fleischhacker, S. E., & Bleich, S. N. (2020). Feeding Low-Income Children during the Covid-19 Pandemic. *New England Journal of Medicine*, 382(18). doi:10.1056/nejmp2005638
- Merrifield, Nicola. "How Covid-19 Exposed Existing Ethnic Inequalities: Socioeconomic Factors and Discrimination - and Not Simply Genetics - May Be behind Higher Death Rates of Non-White People, Including GPs." *Pulse*, July 2020, pp. 14–16. EBSCOhost, search.ebscohost.com/login.aspx?direct=true&db=asn&AN=144681193&site=ehost-live.