Final Project (Part 1 of 3)

Vincent Zhang

Due 2/28 @ 11:59pm

Overview

This is part 1 of 3 of your final project for SSI II. The overarching aims of the project are twofold:

(1) To reflect on and apply some of the concepts and techniques that we have examined this quarter
(2) To develop a basic framework that you can build upon in the spring quarter

Note that it is not necessary to have data on hand now. But you will need to locate some data shortly after submitting this assignment, as part 2 will require you to conduct some initial data visualization and analysis.

At the bottom of this .Rmd, I have listed some common sources of credible data.

Research Question

- State your research question clearly and concisely.
  - Your question should imply a *relationship* between two social concepts. (e.g., Think about the studies we have read.)
- Explain why readers should care about your question.

Currently the "overarching" question I have is a relationship between some defining quality of (demographic) people, and their vaccination rate / hesitancy. For example, here are three relationships we can ask (based on data I took inspiration found from https://www.openicpsr.org/openicpsr/search/covid19/studies?start=0&ARCHIVE=covid19&sort=score%20desc%2CDATEUPDATED%20desc&rows=25&q=vaccine): 1. Is there a significant relationship between the age of a person, and their vaccine hesitancy? 2. Is there a significant relationship between the race of a person and their vaccine rate? 3. Is there a significant relationship between the people with disabilities compared to people without, and their vaccine rates / hesitancy? 4. Is there a significant relationship between the social-economic background of a person, and their vaccine rates / hesitancy?

Here, we see that each of these questions is mainly concerned with a how a specific group of people may result in some kind of higher or lower vaccine rates or vaccine hesitancy than the complement of that group. Notice that I make an intentional seperation between vaccine rates and vaccine hesitancy. Vaccine hesitancy is a measure of how willing these people are to getting the vaccine, which *may*, but not necessarily result in, increased vaccine rates. Vaccine rates is a measure of how many of these people got the vaccine, which *may*, but not necessarily imply that all these people wanted to get the vaccine. These two concepts are related, but they are not the same, for we can also imagine that social or economic barriesr are preventing people who want vaccine from getting one.

Also, I am still unsure which of these question I will actually be using, or whether it is fair to use an umbrella social question of all of these (i.e., is there a relationship between the demographic of a person, and their vaccine hesitancy or vaccine rate)? Additionally, I am not sure whether I want to dive into vaccine rate or vaccine hesitancy, or if it is fair to look at both as a part of my research question. (Any tips here would be appreciated!)

For the purpose of this part 1 assignment, however, I will assume that we are looking at question 1–that is, the relationship between age and vaccine hesitancy (vaccine rate doesn't make as much here, since young children have less control over their own choices)–and analyzing this research question.

Why should we care about this question? Remember that our question should be of social and societal relevance. The answer to our reserach question could be meaningful to knowing more about how our social world works, gain insight to some underlying problem in it, and/or allowing us to address some sort of unfairness, and I propose that answering our research question will allow us to do that. Perhaps our goal is to increase vaccine rates, and one way to get there is to improve vaccine hesitancy (willingness). However, when we make social programs to address, this, we first need to ask ourselves: who are our target audience? Would our program be efficient if it is advertising to people who are already willing to get the vaccine and already have it? In a way, understanding more about the social relationships in the context of the demographics of the people and their vaccine rate/willingness will allow us to more effectively understand and bring about change to the social issues we see.

Hypothesis

- What sort of relationship do expect to find between the concepts in your research question?
- Lay out your logic: Why do you expect the concepts to be linked?

My impression (although this could certainly be wrong), is that younger people are more willing to accept vaccines as a legitamate "science", while the older someone is, the more likely they would believe that it is a "hoax", or that it is too early in the stage to trust it. This is not entirely based, but rather just some inner stereotyping. As a result here is my alternative hypothesis:

The age of a person and their vaccine hesitancy is positively correlated, meaning the older someone is, the more hesitant they are.

My logic for why I could expect this concept is that in modern years, education has become a lot more centered and a part of a children's focus, and as a result younger people are more likely to be more educated on scientific methods, and as a result have less doubts about the effectiveness of vaccine, and thus feel less reason to be against it or believe it is false.

Measurement

- Define the concepts in your research question.
- How might you measure the concepts? Why would you measure them that way?
  - e.g., why measure democracy as elections instead of democracy as elections plus basic civil rights?
  - e.g., why use a dichotomous measure of democracy (0 and 1) instead of an ordinal measure (-10 to +10)?
- Are your measures likely to derive from an experiment, or via observation?

There are two concepts we want to define the measurement of: 1. age 2. vaccine hesitancy

Age is pretty obvious, just measure it as the literal age of the participant. One question we might want to consider though, is if we would want to group the age levels. To gain some intuition on why this would make

sense, consider the imagine that we expect a linear relationship between the increase in age and increase in hesitancy. However, what about the ages between 0 - 5, the members in this group do not even have the cognitive credibility to have valid hestiancy scores. Then, should they be included in our linear graph? As a result, maybe is better that we group the ages by what they classifiy as, such as teen, young adult, middle-age, and old-age etc. This would also provide us a general trend, but also allow us to classify what group each person belongs to. Since I have not yet collected the data, I do not know if there is worthwhile to group the ages, so for now, either measurement (raw ages, and grouped ages) is fine.

Vaccine hesitancy could be on an ordinal scale. The answer for this mainly because hesitancy is something that could only be measured by surveying, in my opinion. An experiment does not make sense, for it is likely to expensive, or not worthwhile, to measure a hypothetical hesitancy on whether they would get a vaccine, when we could just directly survey them. We could develop a scale based on survey questions that allow us to calculate the hesitancy. For example, one what we can acheive an ordinal scale is to literally ask participants how they feel about vaccine, and on a scale of 1-10, how hesitant they are to getting one. This however is subjective, and there could be the question of whether this is a good scale to base our analysis on. Another suggestion is to have participants answer many survey questions that we have already put on a scale from -10 to 10 (based on how strongly it represents being hesitant or for), then averaging out those answers to provide some continuous scale metric that is negative or positive based on whether our participant leans more towards being in favor, or hesitant against vaccines. For example, one such question could be, answer agree, somewhat agree, neutral, somewhat disagree, or disagree to the following: the covid vaccine is safe. If some person answer agree, then it would shift our continous scale more towards the negative, and if some person answers somewhat disgree it would shift our continuous scale slightly to the positive side. Averaging many of these relevant questions would allow us to average out whatt the general opinion that participant has towards vaccines and if they are hesitant torwards it.

I kind of implied it in my answers before, but our measurements should be conducted using a survey (observational instead of experimental.

Data

- It is not necessary to have data on hand now. But you will need to locate some data shortly after submitting this assignment, as part 2 of the final project requires you to perform some initial data analysis and visualization.
- If you have located data, briefly describe the data:
  - Where do the data come from?
  - Which variables are you likely to use in your analysis?

I located my data from this site: https://www.openicpsr.org/openicpsr/project/130422/version/V3/view?path=/openicpsr/130422/fcr:versions/V3/_codebook.xlsx&type=file

This data comes from the following paper: Zhang, Felicia, Shu-Fang Shih, Harapan Harapan, Yogambigai Rajamoorthy, Hao-Yuan Chang, Awnish Singh, Yihan Lu, and Abram L. Wagner. "Changes in COVID-19 Risk Perceptions: Methods of an Internet Survey Conducted in Six Countries." BMC Research Notes 14, no. 1 (December 2021). https://doi.org/10.1186/s13104-021-05846-8.

The variables that I would likely use is: 1. d_age_cat (age in categories) 2. d_age_yr (age in absolute) 3. v (would you accept a coronavirus vaccine?) 4. vax_hes (this is a vaccine hesitancy scale developed from Akel et al. Modification of a vaccine hesitancy scale for use in adult vaccinations in the United States and China) 5. jj_concern (how concerend are you about side effects after receiving the COVID vaccine?)

These variables will allow us to measure both our IV and DV, or in other words, allow us to measure the vaccine hesitancy for each age group we are concerned over.

Analysis Plan

- What will you do with your data in order to assess your research question?
  - What visualizations will you create?
  - What sort of analysis will you conduct and why (e.g., t-test, multivariate regression)?
  - Have you thought about potential confounders to control for? Which ones? Why are they confounders?

Currently, there are two visualizations I am considering. Our data is rather simple currently, in the sense that both our IV and DV are numerical and single varialbe. This leans itself heavily torwards a linear regressional model, but also if we do the categorical measurements on the age groups, then perhaps a geom_bar histogram would also provide valuable inside.

I do not think there is much confounders to be concerned over, although there could be things we might want to take into consideration; confounders are variables that affect IV and DV, and since we our IV is age, there is not much that would literally affect that. One thing that we do have to consider though, is whether our participant has a parent. It is no doubt that a parent can play a strong role on your own opinion on vaccines. Older people may not have parents anymore, or may not be in touch with them, and perhaps it would be different for young people. Thus, parents is a possible confounder because it possibily affects both age of our participant and the opinions they have, and is something we might want to consider.

Data Sources

NOTE: Google's dataset search allows you to search for datasets using keyword search.

- Harvard Dataverse: Repository for data and replication files across disciplines.
- Data.gov: Repository for all US government open datasets.
- Covid 19 Repository: Repository for Covid 19-related data.
- World Bank: Measures of economic and human development for countries across time.
- Pew Research Center: Host of survey data (US and global).
- American National Election Studies: Electoral survey data for US.
- Comparative Study of Electoral Systems: Electoral survey data across countries (multiple waves).
- World Values Survey Survey of social, political, economic, religious and cultural values of people across countries.
- Child & Family Data Archive: Repository for datasets about young children, families, and communities.
- FBI: Data from the FBI on crime across US.

There are many more. The Library at the University of Chicago has additional sources at the following guide: (https://guides.lib.uchicago.edu/data).