

# Final Assignment

Ali Raeisdanaei

December 7, 2022

## 1 Question 1

## 2 Question 2

### 2.1 Part A

A randomised experiment is one where the researchers treat groups differently based on an artificially designed randomness. Whereas the natural experiment relies on events occurring in the real, natural world as the random treatment. Natural experiments are then obviously less intrusive.

### 2.2 Part B

The motivation of this experiment is to study the importance of voting for liberals and conservatives.

A base hypothesis: voter turnout will be lower on days with bad weather, rain, snow, .... Some ridings in Canada have the same candidates for liberals and conservatives. The people of closely neighbouring ridings, have similar demographics and vote about the same. Demographic information, voter turnouts, vote outcomes, and the weather of a given riding are all public sources of data.

To measure the importance of voting for different political sides, we can measure the votes for conservatives and liberals on good and bad weather days.

Of course, there are a lot of confounders. However, there can be many points that control for this. Ridings that have had similar outcomes over the years can be considered to have more weight. The ratio of voter turnout to the vote outcome can be measured for these ridings. Ridings can be compared with during different days of the election. Ridings can be measured with neighbouring ridings with similar political leanings but different weathers. The outcome of the general elections can be used to control and cancel out any changes in political leanings.

### 2.3 Part C

The counting things approach would simply count the number of conservative and liberal voters in comparison with the voter turnout. Ridings of lower turnouts may lean towards one election outcome, conversely ridings of higher turnouts may lean towards a different outcome.

### 2.4 Part D

The second approach is a lot simpler and has fewer confounding factors. The natural experiment seeks to measure differences of turnouts due to weather to correlate with certain outcomes. The second approach simply measures the outcome given the overall turnout. I would set up the research with a layered approach; both methods are useful and would give stronger evidences to the conclusions.

## 3 Question 3

### 1. Big

The paper on Social Capital [1] uses the largest data of 21 Billion friendships. With their large dataset, chetty *et al.* were able to have their results at the granularity level of ZIP Code across the continental United States.

2. Always On

Hangarten *et al.* [3] were able to create a platform for the entire country of Switzerland. This platform functioned as a hiring tool for employers and job-seekers. In this way, they are able to collect the data continuously and study its changes through time.

3. Non-Reactive

Koenecke's article on racial disparities in automated speech is the least reactive [5]. Due both to the size of its data and its nature, the study is not going to affect the results. The research studied the disparities of accuracies of automated speech tools given different dialects and accents of speech in American English. Because of non-reaction, the researchers could then suggest ways to mitigate the racial disparities found.

4. Incomplete

The article of Hangarten [3] is the most incomplete. This project measured the time a user's profile was viewed and the results of the contact button being clicked. This is a large operationalisation and assumption. The data does not have explicit information about any discrimination happening, merely other secondary indicators.

5. Inaccessible

Government arrest records may not always be public; therefore, Kleinburg's paper [4] to focus on New York arrest records. This simplifies the research and the predictive model.

6. Non-representative

Chetty's social capital paper [1] is the data of all social capital with regards to Facebook. This is a huge assumption, and the external validity depends on who uses Facebook. However, they justify their data by the sheer size of population considered.

7. Drifting

Facebook's user base, algorithms, and friendship connections change and drift all the time. Chetty's social capital paper [1] had to deal with drifting of the data over time. The researchers are generally restricted to a snapshot: their results may not hold longitudinally.

8. Algorithmic Confounding

Again Chetty's social capital paper [1] is the most algorithmically confounding. This is one of their biggest threats to validity, as the friendships made by the users under question could be motivated by many unknown algorithmic pushes over time.

9. Dirty

Once again, Chetty's social capital paper [1] is the most dirty. The other datasets from the other articles do not include as many bots as this dataset. A single valid user may be duplicated many times under pseudonyms, spam, fake, and other dirty accounts are very hard to detect. This is a huge challenge and a big threat to validity for this research.

10. Sensitive

While the paper on machine predictions [4] concerns itself with arrests, this information may not necessarily be sensitive. As for the most part, this information is often public, even the court hearings allow for public audiences. However, the data produced by Chetty in the social paper [1] is at the granularity of ZIP code. This information is coupled with social capital and other sensitive details. Such information could be used to identify individuals, groups, or demographics for malicious intents, such as running scams. Such sensitivity is an ethical issue for this research.

## 4 Question 4

The emotional contagion paper ran an experiment of showing more or less positive and negative posts to users [2]. One of the limitations of their experiment, among many, was their classification of positive and negative posts. They simply measure this by the inclusion of positive or negative words in a given text. This simplistic approach could be really inaccurate.

Human computation could have been used to better accurately classify these posts, and improve the ethical implications of the paper. Willing participants could have classified given public posts with sensitive information removed. A given post should be classified multiple times to increase the accuracy.

With this more accurate measure of positivity of text, we can determine a measure of the ratio of positive to negative posts that a user has seen on their timeline. With this ratio, the positivity of posts from the users can also be observed in a more natural way.

This method is observational and less intrusive, so it solves a lot of the ethical issues. The only problem is the private posts being read by the participants. However, many of these posts are on public pages, and the users who are friends with the posters can be used as the participants. So the participants who would see these posts could themselves classify the positivity of the posts. Perhaps a slider is placed below random posts to classify its positivity.

## 5 Question 5

### References

- [1] Raj Chetty et al. “Social Capital I: Measurement and associations with economic mobility”. In: (2022). DOI: 10.3386/w30313.
- [2] “Experimental evidence of massive-scale emotional contagion through social networks”. In: (). DOI: 10.1073/pnas.1320040111.
- [3] Dominik Hangartner, Daniel Kopp, and Michael Siegenthaler. “Monitoring hiring discrimination through online recruitment platforms”. In: *Nature* 589.7843 (2021), pp. 572–576. DOI: 10.1038/s41586-020-03136-0.
- [4] Jon Kleinberg et al. “Human decisions and machine predictions”. In: (2017). DOI: 10.3386/w23180.
- [5] Allison Koenecke et al. “Racial disparities in automated speech recognition”. In: *Proceedings of the National Academy of Sciences* 117.14 (2020), pp. 7684–7689. DOI: 10.1073/pnas.1915768117.