Tutorial 6

INF312: Worlds Become Data

Luca Carnegie

February 12, 2024

With reference to Dean (2022), please discuss the difference between probability and non-probability sampling. Please write at least two pages.

When sampling a population, two ways are often used - probability and non-probability sampling. However, each genre comes with their own characteristics that can affect the data collected for a study. Therefore, it is important to notice and understand their impacts to be more critical, effective consumers and producers of knowledge. We will use this understanding to understand Dean (2022)'s paper about the effect on tracking methods on government COVID-19 management.

Non-Probability Sampling

Non-probability sampling is a technique where the selection of samples is based on non-random methods, meaning that all members of the population do not have an equal chance of selection for the sample.

There are a few different sub-types of non-probability sampling:

- **Convenience Sampling:** Samples are chosen because they are simply easy to access for the researcher.
- **Purposive sampling:** Samples are chose on the researcher's judgement that that they are representative of the population of interest.
- **Quota sampling:** Quotas are set for different subgroups of the population and then samples are then chosen until the quotas are filled. Subgroups would be designated based on a particular set of characteristics.
- **Snowball sampling:** involves starting with a small group of samples and then asking them to identify other potential samples who meet the criteria until a sufficient sample size is reached.

Non-probability sampling is sometimes used in quantitative research, especially when it is not possible to obtain random samples of the population. Consequently, though, non-probability sampling does not allow you to make generalizations about the population from your sample. This is because you cannot be sure that your sample is representative of the population.

Probability Sampling

Probability sampling is where each member of a population has a **known and equal probability of being selected** for inclusion in a sample. Probability sampling is essential for ensuring that the results of a study can be generalized to the wider population, as it avoids bias and allows for the calculation of representative statistics. It is the sampling method that what advocates for more of during the fight against COVID-19, as she discusses in her article.

There are a few different sub-types of probability sampling:

- **Simple random sampling:** Each member of the population has an equal chance of being selected. This is often done by assigning numbers to all members and then randomly selecting those numbers.
- **Stratified sampling:** The population is divided into subgroups (strata) based on shared characteristics, and then a random sample is drawn from each stratum. This ensures that all subgroups are represented in the sample.
- **Cluster sampling:** The population is divided into groups (clusters), and then a random sample of clusters is selected. All members of the selected clusters are then included in the sample.

However, probability sampling comes with challenges. One is the difficulty of obtaining a truly representative sample, especially in populations that are difficult to reach or where participation rates are low. Additionally, random sampling can be time-consuming and expensive to implement, particularly in large-scale studies. In the case of Dean (2022), excessive cost appears to be a barrier to government investment, hence Dean's advocacy for it in her article.

Discussion

Now that the definitions of probability and non-probability sampling methods, discussing Dean (2022) becomes somewhat simpler. In her article, Dean asserts that the various data collection methods used by epidemiologists are not very good at understanding the actual changes happening with infection rates. Often, these methods are examples of using non-probability sampling, which cannot be used to generalize a population, but are used to do just that. This, she argues, had led to sometimes extreme information asymmetry with respect to the costs and benefits of enacting a certain government policy.

Random sampling can help allievate this information assymetry. As Dean says: "as long as participants are selected randomly, they will on average mimic characteristics of the wider population. Roughly speaking, testing fewer than 1,000 people can yield crucial information about 10 million, or even more."

Overall, with the now-normal presence of COVID-19 as a common pathogen, investing in most costly random sampling can better prepare governments for future pandemics and allow them to make faster decisions that can save more lives.