Estimating Statewide Respondent Totals Using the Laplace Ratio Estimator: A Case Study with 2022 ACS Data on Doctoral Degree Holders*

Arav Agarwal Daniel Du Dong Jun Yoon Grace Nguyen

October 3, 2024

Data

This repository uses data from IPUMS (*IPUMS USA* (2022)), and is analyzed with R (R Core Team (2023)). To obtain the data, follow these steps:

1. Google "IPUMS USA" and go to the first link: https://usa.ipums.org/usa/.
2. Click "Get Data" on the home screen.
3. Click "Select Samples". In USA Samples, tick the box that says "Default sample from each year", and then tick the box next to 2022 ACS. This should the only box ticked. Afterwards, click "submit sample selection".
4. Under Select Harmonized Variables, click the Household drop down menu and select Geographic. Click the + under "Add to cart" next to variable `STATEICP`.
5. Do the same for variable `SEX` in Demographic under the Person menu, and for variable `EDUC` in Education under the Person menu.
6. Click "View Cart". You should be taken to the extract request, with 1 sample and 14 variables. Change the data format to .csv, and then press "Submit Extract".
7. If you are not logged in to an account, you will be prompted to log in, or create an account. Once you are finished this, you can view and download the csv extract. There is also a link to the codebook that explains the variables in the dataset.

By using the codebook, we filter the amount of respondents in each state `STATEICP` that had a doctoral degree in `EDUCD` as their highest educational attainment. Here are the results in a table:

*Code and data are available at: IPUMS USA.

Overview of ratio estimators approach

Background

The purpose of the ratio estimator is to estimate a population parameter (such as a total or mean) using auxiliary information that is correlated with the variable of interest. By leveraging this auxiliary variable, the ratio estimator improves the precision of the estimate compared to simpler methods, like the sample mean. Rather than directly estimating the total number of respondents across all states, the ratio estimator uses a known ratio between two quantities—such as the number of people with doctoral degrees and the total number of respondents in California—and applies this ratio to other states. This method assumes that the relationship observed in California is approximately true for other states.

Approach

We are given the total number of respondents in California, 391171, as well as the total doctoral respondents in each state. To estimate the total number of respondents in each state we first calculate the following ratio:

Number of Respondents in a State : Number of Doctoral Respondents

That is, Ratio = 391171/n_Doctoral_CA

Then we extrapolate using this ratio by applying this ratio to our column of the number of respondents with a doctoral degree in each state. That is, we multiply each value in the collumn with the ratio we calculated.

Estimating Total Respondents

If, for example, there are 391,171 respondents in California (state 71 according the codebook) across all levels of education, we can estimate the total number of respondents in each state as follows:

California respondents: 391171

California respondents with a doctorate: 6336

Ratio: 391171/6336 = 61.738

The estimates are likely different because of the following data limitations, discussed below:

Homogeneity Assumption:

The biggest assumption here is that the ratio of doctoral degree holders to total respondents is consistent across all states, which might not be true due to various socio-economic, educational, and cultural differences.

Variability in Smaller Populations:

States with very small populations or low numbers of doctoral degree holders might see more variability and less accuracy in the estimates.

References

IPUMS USA. 2022. https://usa.ipums.org/usa/.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.