Estimating U.S. State Populations Using the Ratio Estimator Approach with IPUMS Data*

Qingyang Feng

October 3, 2024

Introduction

To obtain the necessary data from the IPUMS USA website, start by accessing the "Select Data" section. First, deselect the default options and choose only the 2022 ACS dataset. Then, under the Households section, navigate to GEOGRAPHIC and select the variable STATEICP. Next, under the Person section, go to EDUCATION and choose EDUC, and under DEMOGRAPHIC, select SEX. After selecting all the relevant variables, submit your data extract request. If you do not already have an account, you will need to create one. Once the request is processed, download the dataset in CSV format. By following these steps, you will successfully obtain the required data for analysis.

Overview

The ratio estimator approach is a statistical technique used to estimate unknown population totals by leveraging known ratios. In this document, the approach involves dividing the number of respondents with a doctoral degree in California by the total number of respondents in California, yielding a ratio of approximately 0.0162. This ratio is then applied to other states to estimate their total populations based on the number of doctoral degree holders, assuming a similar proportionality across states. By using this method, we enhance the accuracy of estimates by utilizing known, existing data.

To implement this, we first group the data by STATEICP, which is the state identifier, and count the number of doctoral degree holders in each state. We then use the California ratio to estimate the total number of respondents in each state by dividing the number of doctoral

*Code and data are available at: https://github.com/Fqy10987/US_Doctorate_Analysis.git.

degree holders by the ratio. This assumes a consistent proportion of doctoral degree holders across all states, thereby allowing us to estimate each state's total respondent count.

Result

```
# A tibble: 51 x 2
   STATEICP doctoral_count
      <dbl>          <int>
 1        1            600
 2        2            165
 3        3           2014
 4        4            244
 5        5            177
 6        6            131
 7       11            152
 8       12           1438
 9       13           2829
10       14           1620
# i 41 more rows
```

```
# A tibble: 51 x 3
   STATEICP actual_total estimated_total
      <dbl>        <int>           <dbl>
 1        1        37369          37043.
 2        2        14523          10187.
 3        3        73077         124340.
 4        4        14077          15064.
 5        5        10401          10928.
 6        6         6860           8088.
 7       11         9641           9384.
 8       12        93166          88779.
 9       13       203891         174656.
10       14       132605         100015.
# i 41 more rows
```

Some explanation of why you think they are different

The estimated total number of respondents in each state using the ratio estimator approach may differ from the actual number for a few reasons:

Different Proportions: The method assumes that the proportion of people with doctoral degrees in California is similar to other states, but that's not always the case. Education levels can vary a lot depending on the demographics, job opportunities, and education systems in each state, which means the estimates might not always match the reality.

Sampling Differences: If the data we're using is just a sample and not the whole population, there's a chance that the sample doesn't fully reflect what's happening in the entire population, which can throw off the estimates.

Uneven Education Levels: Educational attainment isn't the same everywhere. Some regions have more access to higher education or different policies that encourage people to pursue advanced degrees, so it's unlikely that California's ratio of doctoral degree holders applies evenly across all states.

Potential Bias: The ratio estimator works well if the relationship between the number of doctoral degree holders and the total population is similar in every state. But if California's ratio isn't a good reflection of other states, the estimates will be biased and less accurate.