Estimating Statewise Respondants in the 2022 ACS

Viet Nguyen, Yihang Xu, Doran Wang

Introduction

With a particular focus on people whose greatest educational attainment is a doctorate, the goal of this analysis is to estimate the total number of respondents in each U.S. state from the 2022 American Community Survey (ACS) using a ratio estimator approach. Based on a sample of respondents from each state, the ACS offers insightful information about the social, economic, housing, and demographic aspects of the American people. However, we estimate these totals using a ratio-based method, rather than having the exact respondent counts available for all states.

The ratio estimator technique implies that there is a consistent link between the number of individuals with PhD degrees and the total number of respondents across different states. We compute a ratio based on the number of people in California who hold doctorates, using the actual number of respondents (391,171) as a point of reference. We then use this ratio to calculate the total number of responders in other states by taking into account the number of people who hold doctorates in each state.

Instructions on obtaining data

1. Create an account on IPUMS USA
2. Select the 2022 ACS dataset in sample selections
3. In the select variables, these are included:
    - STATEICP
    - SEX
    - EDUC
4. Submit data extract, then download the csv file once the extract is processed.

Ratio Estimators Approach

We know that the total number of respondents in California is given by:

$N_{\text{cal}} = 391171$

Let D_{cal} be the number of respondents in California with a doctoral degree. The ratio for California is:

$r = \frac{N_{\text{cal}}}{D_{\text{cal}}}$

To estimate the total number of respondents for any other state i, we use the number of doctoral degree holders D_i and apply the ratio:

$\hat{N}_i = r \times D_i$

Where: - $\hat{N}_i$ is the estimated total number of respondents in state i, - D_i is the number of doctoral degree holders in state i.

Running Code

```
# A tibble: 51 x 3
   STATEICP doctoral_count estimated_total
      <int>          <int>           <dbl>
 1        1            600          37043.
 2        2            165          10187.
 3        3           2014         124340.
 4        4            244          15064.
 5        5            177          10928.
 6        6            131           8088.
 7       11            152           9384.
 8       12           1438          88779.
 9       13           2829         174656.
10       14           1620         100015.
# i 41 more rows
```

```
# A tibble: 6 x 2
  STATEICP actual_total
     <int>        <int>
1        1        37369
2        2        14523
3        3        73077
4        4        14077
5        5        10401
6        6         6860
```

Discussions

Although the ratio estimator offers a handy method for estimating population sizes, it is predicated on the idea that the proportion of Californians with doctorates to the state's overall population is typical of other states. This assumption does not always hold true in practice, as evidenced by variations in state-specific legislation, educational systems, and demography. As a result, there are disparities between the estimated and actual number of respondents among states.

These variances underline the need of addressing geographical variations and employing more advanced modeling methodologies when producing population estimates based on specific demographic categories.