

An Analysis of 2022 ACS Data from IPUMS

Ricky Yuan

Using the data from 2022 ACS, showing the respondents in each state of STATEICP that had a doctoral degree as their highest educational attainment of EDUC.

```
# 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
```

Instructions on how to obtain the data.

The data from the website of IMPUS of USA, and the I creates the account that can help me to download the data. After log in, selecting data and click on “SELECT DATA” in the top navigation bar to access the data selection page. In the Sample Selection section, select data for the 2022 ACS. Next, click on “SELECT VARIABLES” on the left side of the page. In the search bar, look for “STATEICP” (for state) and add it to the data cart, then look for “EDUCD” (for educational attainment). After selecting your variables, click View Cart at the top right of the page and then click Create Data Extract. Next, name the extract and submit the request, which will be processed in the background. Check the status in My Data Extracts when the extraction is complete. Finally, download the .csv file.

A brief overview of the ratio estimators approach.

The ratio estimators approach uses the proportion of respondents with doctoral degrees in California to estimate the total number of respondents in each state across the U.S. Specifically, we calculate the ratio of doctoral degree holders to the total respondents in California and assume this ratio is similar across other states. By applying this ratio to the total population of each state, we can estimate the number of doctoral degree holders of each state. This approach leverages California's data as a baseline to make inferences about the broader distribution of education levels across the country.

Based on the dataset, we find that the total number of respondents in California is 391,171, with 6,336 of them holding doctoral degrees. This results the doctoral degree holders ratio of approximately 0.01619752.

Your estimates and the actual number of respondents.

```
# 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.

It is biased to assume that the proportion in one state (e.g., California) is representative of other states. The reasons for the bias in the projections are as follows: The main reasons for differences between estimated and actual respondent numbers include the similarity assumption, which presumes uniformity across states in the ratio of doctoral degree holders. However, each state has unique demographics, economic conditions, and educational policies. States with more universities and research institutions naturally have more doctoral degree holders compared to rural or economically less developed states. Additionally, the non-uniform

distribution of advanced degrees is evident: regions dominated by high-skill industries like technology, healthcare, or academia (e.g., California's Silicon Valley or Boston's healthcare hubs) have higher demand for and attainment of advanced degrees. These areas offer favorable job markets for advanced degree holders, attracting more individuals pursuing such qualifications. In contrast, states with less advanced economies show significant differences in degree distribution. Extrapolating national estimates based on high-tech regions can distort results, misrepresenting rural or non-high-tech areas. Lastly, sampling bias may occur if the sample data does not fully represent the population in certain states, leading to inaccurate estimates of educational attainment.