

Estimating State-Level Doctoral Degree Attainment Using the 2022 ACS*

Yuanyi (Leo) Liu Xuanle Zhou Dezhen Chen Yongqi Liu
Ziyuan Shen

October 3, 2024

This document analyzes the number of respondents in each state who have a doctoral degree as their highest educational attainment in the 2022 ACS IPUMS data. We estimate the total number of respondents in each state using the ratio estimators approach based on data from California.

Table of contents

1	Instructions on how to obtain the data	1
2	A brief overview of the ratio estimators approach	2
3	Estimates and the Actual Number of Respondents	2
4	Some explanation of why you think they are different	2

1 Instructions on how to obtain the data

To extract and download data from [IPUMS USA](#), began by selecting “Get Data” and proceeded to “Select Sample.” Then deselected the “Default sample from each year” option and opted for the “2022 ACS” sample only. After confirming the selection, moved to the “Household” section, chose “Geographic,” and selected “STATEICP.” Similarly, under the “Person” section, selected “sex” and “EDUC.” Next, reviewed variable choices by clicking “View Cart” and then clicked “Create Data Extract.” Set the “Data Format” to “.csv” while keeping the “Data Structure” as “Rectangular.” After submitting the request, logged in the account, waited for

*Code and data are available at: [Estimating State-Level Doctoral Degree Attainment](#).

the email notification that the extract was ready, and downloaded the file. Eventually, unzip the data using the command `gunzip usa_00002.csv.gz` to proceed with the analysis.

2 A brief overview of the ratio estimators approach

The ratio estimators approach is a method used to estimate a demographic parameter such as total or mean which is using the relationship between two relevant variables. For this assignment, we selected the California population group. This group has a specific characteristic such as the proportion of the population with a doctoral degree. Then, we use the percentage of the population with a doctoral degree to estimate unknown characteristics in the general population. We can use this method when we don't know the exact size of the population, but when we can find enough sample data to assume that these proportional relationships are consistent across groups.

3 Estimates and the Actual Number of Respondents

```
# A tibble: 51 x 3
  STATEICP actual_total estimated_total
  <fct>      <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
```

4 Some explanation of why you think they are different

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

- Assumption of Similarity: The ratio estimator operates under the assumption that the percentage of individuals holding doctoral degrees in California mirrors that of other states. However, educational attainment can differ significantly from state to state due to

varying demographics, economic conditions, and educational systems. These differences contribute to inconsistencies between the estimated numbers and the actual figures.

- **Sampling Variability:** If the data used for estimation is a sample data rather than a complete population, random sampling variability will affect the calculated ratio and its accuracy.
- **Uneven Distribution:** Educational attainment is not uniformly distributed across the United States. Regional policies, cultural differences, and access to higher education vary, meaning the ratio in California may not apply to other states.
- **Bias in the Ratio:** The Laplace ratio estimator may be biased if the relationship between doctoral degree holders and the total population in one state, such as California, differs from that in other states due to unobserved factors.

In summary, assuming homogeneity in educational attainment across states when using ratio estimators can lead to differences between the estimates and actual numbers.