Estimating the Number of Doctoral Degree Holders in the US: A Comparative Analysis Using Laplace Ratio Estimators*

Discrepancy Between Estimated and Actual Total Respondents Show Regional Variations in Educational Attainment

Yunkyung Ko YingKe He Richard Guo Ruikang Wang
Ziheng Zhong

November 21, 2024

This paper uses the 2022 American Community Survey (ACS) data from IPUMS USA to estimate the total number of respondents with doctoral degrees across U.S. states using the Laplace ratio estimator. The estimated totals are compared to actual state-level respondent counts, highlighting discrepancies caused by regional variations in educational attainment. The results emphasize the limitations of applying a single ratio across diverse states and suggest that regional differences should be accounted for when using such estimation methods.

1 Method of Obtaining the Data

To retrieve data from IPUMS USA (Minnesota Population Center 2022), we first login to the IPUMS website and selected "IPUMS USA." From there, we clicked "Get Data" and chose the "2022 ACS" sample under the "SELECT SAMPLE" section. To obtain state-level data, we selected "HOUSEHOLD" > "GEOGRAPHIC" and added the "STATEICP" variable to the data cart. For individual-level data, we went to the "PERSON" section and included the "EDUC" variable. After reviewing the cart, we clicked "CREATE DATA EXTRACT" and set the "DATA FORMAT" to "csv." Finally, we submitted the extract and then downloaded and saved the file locally (e.g., "usa_00001.csv") for use in R (R Core Team 2023). The `tidyverse` package (Wickham et al. 2019) was used to enhance the efficiency of data manipulation and data visualization

*Code and data are available at: https://github.com/koyunkyung/state_respondents.

Next, we count the respondents with doctoral degrees grouped by states as shown in. Table 1 presents an overview of the distribution of respondents with doctoral degrees across different states in the U.S., which is crucial for the understanding of regional educational attainment trends. The first column, 'State', refers to the unique state or territory codes provided by the IPUMS system, where each number corresponds to a specific state. The second column, 'Respondents with Doctoral Degree', provides the number of respondents with doctoral degrees (filtered by EDUCD == 116) in each state. For example, state code 1 shows 600 respondents with doctoral degrees, while state code 3 shows a significantly higher number with 2014 respondents. Other state codes show varying counts of respondents, such as 165 for state code 2, 244 for state code 4, and so on.

Table 1: The distribution of respondents with doctoral degrees across states with different STATEICPs (**Note:** The table is being displayed with only the first part shown to optimize space.)

State	Respondents with Doctoral Degree
1	600
2	165
3	2014
4	244
5	177
6	131

2 Brief Overview of the Ratio Estimators Approach

The ratio estimator is a statistical method used to estimate population totals or averages by utilizing known sample ratios. It operates by calculating the ratio of a specific characteristic (in this paper, the number of individuals holding doctoral degrees) relative to the total population in a known sample (the state of California). This ratio is then applied to other regions or groups, based on the assumption that the relationship between the characteristic and the population remains consistent across different areas. This method is particularly effective when complete population data is unavailable, but sample data provides a reliable basis for inferring proportional relationships that can be generalized.

3 Estimates and the Actual Number of Respondents

When assuming that California has 391,171 respondents across all educational levels, we can calculate the ratio for California. The Laplace ratio method works well when the relationship between the characteristic of interest, such as the proportion of doctoral degree holders, and

the population is consistent across all units. In the example below, the ratio for California is 0.01619752.

```
[1] "The California ratio that was found: 0.0162"
```

Using the ratio estimator approach estimates the total number of respondents for each state as shown in Table 2. Assuming the ratio holds for all states, we can estimate the total respondents for every state.

Table 2: Total number of respondents in each state, estimated by using the ratio estimators approach (**Note:** The table is being displayed with only the first part shown to optimize space.)

State	Respondents with Doctoral Degree	Estimated Total Number of Respondents
1	600	37037.04
2	165	10185.19
3	2014	124320.99
4	244	15061.73
5	177	10925.93
6	131	8086.42
11	152	9382.72
12	1438	88765.43
13	2829	174629.63
14	1620	100000.00

Table 3 provides a comparison between the estimated total respondents, derived using the ratio estimator method, and the actual total respondents for each state. The 'Respondents with Doctoral Degree' column represents the number of individuals with doctoral degrees in each state, while the 'Estimated Total Respondents' column shows the projected total population for each state, calculated by applying the California ratio (0.01619752) to the number of doctoral respondents in each state. The 'Actual Total Respondent' column reflects the true population count for each state, which was directly obtained from the data. This allows for a comparison between the estimated and actual figures. In state code 1, the actual number of respondents is 37,369, slightly higher than the estimate of 37,042. For state code 3, the actual total is 73,077, significantly lower than the estimate of 124,340. These differences illustrate the potential inaccuracies that can arise from using a ratio estimator when the relationship between the doctoral respondents and the total population is not consistent across states. This highlights the importance of accounting for regional variations in such analyses.

Table 3: The comparison of actual and estimated total number of respondents in each state (**Note:** The table is being displayed with only the first part shown to optimize space.)

State	Respondents with Doctoral Degree	Estimated Total Respondents	Actual Total Respondents
1	600	37037.04	37369
2	165	10185.19	14523
3	2014	124320.99	73077
4	244	15061.73	14077
5	177	10925.93	10401
6	131	8086.42	6860

Figure 1 uses a bar chart to show the differences between estimated and actual total respondents across every state of the U.S.

4 Explanation of Why They are Different

A simple explanation for why the estimated number of respondents and actual number of respondents differ is that the naive assumption that all states share the same doctorate-respondent to total respondent ratio is false. The ratio estimator assumes that the proportion of respondents with doctoral degrees in California is representative of the proportion in other states. However, educational attainment can vary significantly due to differences in demographics, economic opportunities, and educational infrastructure across states. This variance leads to discrepancies between the estimated and actual counts. When the data used for estimation is based on a sample rather than a complete population census, random sampling variability introduces uncertainty into the calculated ratio, which in turn affects the accuracy of the estimates. This variability can lead to deviations between the sample-based ratio and the true population ratio, impacting the reliability of the estimation.

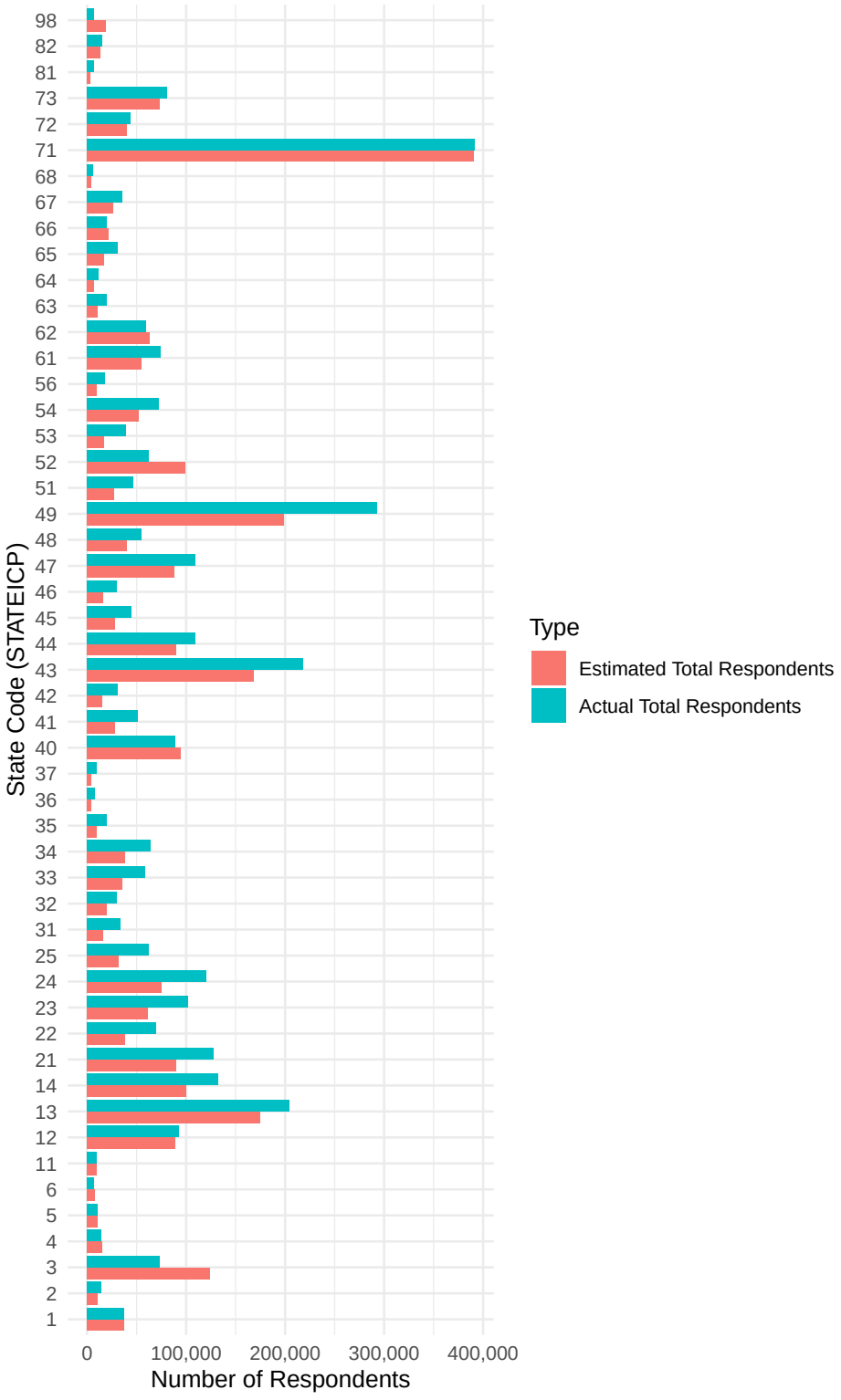

Figure 1: Comparison of estimated and actual total respondents in each state of U.S.

References

Minnesota Population Center. 2022. "IPUMS USA: Version 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/.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.