A Case Study from 2022 IPUMS USA*

Analyzing state-level populations using the ratio estimation method reveals critical differences in population distribution.

Yuxuan Wei

November 21, 2024

Introduction

This study utilizes the 2022 American Community Survey (ACS) dataset, obtained from the Integrated Public Use Microdata Series USA (IPUMS 2024). The primary goal is to analyze the relationship between educational attainment, specifically doctoral degrees, and state-level population estimates using the **ratio estimation method**.

The ratio estimation method provides a way to generalize known relationships within a sample (e.g., California) to estimate populations across states. By understanding the proportion of doctoral degree holders in California and applying this to other states, we estimate total populations. This approach reveals both the utility and limitations of ratio estimators in population studies.

Data are extract from IPUMS_USA. Two variables were extracted from the 2022 ACS: STATEICP and EDUC. After selecting the correct data and converting the file from a .dat to a .dta, you could download the data extract from IPUMS USA. You then unzip the file and import it into R (R Core Team 2023).This analysis uses the dplyr, tidyr, and haven packages in order to synthesize the data (Wickham et al. 2019).

Doctoral Degree Holders by State

In Table 1, we observe the number of doctoral degree holders by state. For example, state code 3 (California) has the highest number of doctoral degree holders with 2014, while states like 2 (Massachusetts) and 6 (Indiana) have significantly fewer, with 165 and 131, respectively. These differences highlight the uneven distribution of doctoral degree holders across states.

---

*Code and data are available at: https://github.com/wyx827/IPUMS_EDUC

The variation in these numbers reflects the unique characteristics of each state. For example

- States like California (STATEICP = 3) and Florida (STATEICP = 12) have large populations, leading to a higher count of doctoral degree holders.

- In contrast, smaller states like Indiana (STATEICP = 6) show lower numbers due to their smaller populations or lower rates of educational attainment.

This foundational data allows us to compute ratios, which we later use for estimating the total population in each state.

Table 1: Doctoral Degree Holders by State

State Code	Doctoral Degree Holders
1	600
2	165
3	2014
4	244
5	177
6	131
11	152
12	1438

Ratio Estimation for State Populations

The ratio estimation method calculates a baseline proportion by using California's population data, where both the total number of respondents and the number of doctoral degree holders are known. The ratio is then applied to other states to estimate their total populations based on the number of doctoral degree holders in each state.

In Table 2, we compare the estimated total population with the actual total population for a few states:

- California (State Code = 3), which serves as our baseline, has an estimated total of 124,340 and an actual total of 73,077. This overestimation is likely because the proportion of doctoral degree holders in California is used as a reference, and California may have different characteristics compared to other states.

- Massachusetts (State Code = 2) has an estimated total of 10,187 compared to an actual total of 14,523. This discrepancy indicates that Massachusetts has a different ratio of doctoral degree holders relative to its population than California, leading to an underestimation.

- For New York (State Code = 13), the estimated total (174,656) is lower than the actual total (203,891). This suggests that New York's proportion of doctoral degree holders differs significantly from California's, likely due to regional differences in education levels and demographics.

These examples illustrate that the ratio estimation method is sensitive to the underlying assumptions about the relationship between doctoral degree holders and the general population across states. Significant variations in educational attainment levels between states can lead to discrepancies between estimated and actual populations, emphasizing the need for careful consideration when applying ratio estimations to diverse regions.

Table 2: Estimated vs Actual Population Using Ratio Estimation

State Code	Estimated Total	Actual Total
1	37043	37369
2	10187	14523
3	124340	73077
4	15064	14077
5	10928	10401
6	8088	6860
11	9384	9641
12	88779	93166
13	174656	203891
14	100015	132605

References

IPUMS. 2024. "IPUMS USA." https://usa.ipums.org/usa/#:~:text=IPUMS%20USA%20col lects,%20preserves%20and%20harmonizes.

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.