

# Estimation of Respondents in Each State\*

Wen Han Zhao, Yun Chu, Felix Li, Yanzun Jiang

November 21, 2024

## Table of contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Overview of the Ratio Estimators Approach</b>	<b>2</b>
2.1	Estimation Formula and Explanation . . . . .	2
<b>3</b>	<b>Summary Table of Estimated and Actual Respondents</b>	<b>3</b>
<b>4</b>	<b>Explanation of Differences</b>	<b>4</b>
<b>A</b>	<b>Appendix</b>	<b>5</b>
A.1	Instructions on Obtaining the Data . . . . .	5
	<b>References</b>	<b>6</b>

---

\*Code and data are available at: [<https://github.com/chuyun2024/Estimation-of-Doctors/tree/main>]

# 1 Introduction

The American Community Survey (ACS) conducted by the U.S. Census Bureau provides vital demographic, social, economic, and housing data that supports informed decision-making across various sectors. Using the 2022 ACS dataset accessed through IPUMS, this report delves into state-level variations in educational attainment, with a specific focus on respondents holding doctoral degrees. By analyzing these patterns, we aim to shed light on the distribution of advanced educational qualifications across the United States.

To complement the analysis, a ratio estimator method is applied, leveraging the known respondent total for California to estimate total respondents across other states. This approach not only facilitates comparisons but also provides insights into the broader demographic landscape, enriching our understanding of educational trends and their implications.

This report begins by outlining the methodology used for data extraction and analysis, followed by detailed findings on the prevalence of doctoral degrees across states. The discussion highlights key observations, potential limitations, and the broader significance of these trends for policymaking and resource allocation.

This report uses the IPUMS data (IPUMS USA 2022) and it uses the libraries tidyverse (Wickham et al. 2019) and knitr (Xie 2023) on R (R Core Team 2023) codes. Section 2 contains the overview of the Ratio Estimators Approach. Section 3 contains all the summary table of estimated and actual respondents by state. Section 4 contains a brief explanation of the differences appeared due to methodology used. Lastly, Section A contains information to gather data.

## 2 Overview of the Ratio Estimators Approach

### 2.1 Estimation Formula and Explanation

1. Let:

- $R_{CA}$  be the ratio of the number of people with doctoral degrees to total respondents in CA.

- $N_{doctoraldegrees,state}$  be the number people with of doctoral degrees in any other state.

- $N_{respondents,state}$  be the estimated number of respondents in other states.

2. Application to Other States: For any other state, the estimated number of total respondents can be expressed as:

$$N_{respondents,state} = \frac{N_{doctoraldegrees,state}}{R_{CA}}$$

This uses the ratio observed in California to estimate the total number of respondents based on the number of doctoral degrees in other states.

This approach leverages the assumption that the proportion of doctoral degree holders within the survey population is relatively consistent across states, allowing us to use California as a reference point for estimation. While this simplifies the calculation and provides a practical means of estimating total respondents, it is important to recognize that state-specific factors, such as varying levels of educational attainment, population density, and demographic composition, may cause deviations from this assumed uniformity. Despite these potential limitations, this method is a useful starting point for approximating total respondent numbers in the absence of complete data for all states.

### 3 Summary Table of Estimated and Actual Respondents

Table 1: Estimated and Actual Respondents

state_id	estimated	actual	diff
Connecticut	37043	37369	326
Maine	10187	14523	4336
Massachusetts	124340	73077	-51263
New Hampshire	15064	14077	-987
Rhode Island	10928	10401	-527
Vermont	8088	6860	-1228
Delaware	9384	9641	257
New Jersey	88779	93166	4387
New York	174656	203891	29235
Pennsylvania	100015	132605	32590
Illinois	89952	128046	38094
Indiana	38277	69843	31566
Michigan	61182	101512	40330
Ohio	74888	120666	45778
Wisconsin	31672	61967	30295
Iowa	15928	33586	17658
Kansas	19818	29940	10122
Minnesota	35314	58984	23670
Missouri	38339	64551	26212
Nebraska	9446	19989	10543
North Dakota	3704	8107	4403
South Dakota	4383	9296	4913
Virginia	94521	88761	-5760
Alabama	28399	51580	23181

Table 1: Estimated and Actual Respondents

state_id	estimated	actual	diff
Arkansas	15496	31288	15792
Florida	168606	217799	49193
Georgia	89582	109349	19767
Louisiana	27782	45040	17258
Mississippi	16237	29796	13559
North Carolina	87729	109230	21501
South Carolina	39944	54651	14707
Texas	198549	292919	94370
Kentucky	27659	46605	18946
Maryland	99274	62442	-36832
Oklahoma	17348	39445	22097
Tennessee	51922	72374	20452
West Virginia	9816	18135	8319
Arizona	55317	74153	18836
Colorado	63652	59841	-3811
Idaho	10804	19884	9080
Montana	6976	11116	4140
Nevada	17410	30749	13339
New Mexico	21608	20243	-1365
Utah	26424	35537	9113
Wyoming	4445	5962	1517
California	391171	391171	0
Oregon	39944	43708	3764
Washington	73777	80818	7041
Alaska	3149	6972	3823
Hawaii	13212	14995	1783
District of Columbia	19200	6718	-12482

## 4 Explanation of Differences

If we estimate the total number of respondents using the ratio estimator approach with Laplace’s method, we are assuming that the proportion of people with doctoral degrees is the same across all states. However, this is not necessarily the case. California, with its many large universities and research institutions, likely has a higher concentration of individuals with doctoral degrees. As a result, the doctor-to-respondent ratio in California would be higher than in other states, leading to an underestimate of the total number of respondents in those states.

## A Appendix

### A.1 Instructions on Obtaining the Data

1. Go to the **IPUMS website** and click on “VISIT SITE” under the IPUMS USA section.
2. Click on “**Get Data**”.
3. Click “**Select Samples**”, uncheck “**Default sample from each year**”, and check only “**ACS**” for the year 2022.
4. Click “**Submit Sample Selections**”.
5. Under **HOUSEHOLD**, hover over “**GEOGRAPHIC**” and add “**STATEICP**” to your cart.
6. Under **PERSON**, hover over “**DEMOGRAPHIC**” and add “**SEX**” to your cart.
7. Under **PERSON**, hover over “**EDUCATION**” and add “**EDUC**” to your cart.
8. In your **Data Cart**, you should now see 4 variables and 1 sample. Click “**View Cart**”.
9. Click “**Create Data Extract**”.
10. Change the **Data Format** from **.dat** to **.csv** and click “**Apply Selections**”.
11. Click “**Submit Extract**”.
12. Wait for the status to become **COMPLETED**, then click “**Download .csv**” under the “Download Data” column.
13. Finally, unzip the downloaded file to access your data.

## References

- IPUMS USA. 2022. “American Community Survey (ACS) 2022.” University of Minnesota. <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 Golemund, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Xie, Yihui. 2023. *knitr: A General-Purpose Package for Dynamic Report Generation in R*. <https://yihui.org/knitr/>.