

IPUMS 2022 ACS Data Analysis Using Ratio Estimation*

Estimating State-Level Populations Using Ratio Estimation

Yizhe Chen

Charlie Zhang

Qizhou Xie

November 22, 2024

This study uses the ratio estimation method to analyze state-level population data from the 2022 ACS. By using California's known respondent data as a reference, the method estimates populations in other states based on the ratio of doctoral degree holders. The analysis highlights differences between estimates and actual figures, driven by demographic diversity and sampling variability, offering insights into the strengths and limitations of ratio estimation in demographic research.

Table of contents

1	Introduction	2
2	Instructions on How to Obtain the Data	2
3	Overview of the Ratio Estimators Approach	3
4	Estimates and the Actual Number of Respondents	3
5	Explanation of Differences	4
	References	6

*Code and data are available at: <https://github.com/YizheChenUT/Doctoral-Degree-Analysis.git>

1 Introduction

This analysis explores the use of the ratio estimator method through R language (R Core Team 2023) to evaluate population characteristics based on the 2022 American Community Survey (ACS) data obtained from IPUMS USA (Ruggles et al. 2022). By using detailed state-level data on doctoral degree holders and total respondents, the study aims to estimate population metrics for states across the U.S. The method assumes that the relationship observed in a reference state (California) can be extended to other states to derive meaningful approximations.

The dataset includes variables such as STATEICP (state identifiers) and EDUCD (educational attainment), enabling the calculation of a ratio for doctoral degree holders relative to the total population in California. This ratio is applied to other states, allowing for estimations of their total respondent numbers.

The analysis also highlights discrepancies between estimated and actual values, investigating contributing factors such as demographic diversity, sampling variability, and state-level educational trends. This study provides a foundation for understanding the application of ratio estimators in demographic analysis and underscores the potential challenges when generalizing from a single-state reference point.

2 Instructions on How to Obtain the Data

To obtain the IPUMS 2022 ACS data (Ruggles et al. 2022):

- 1. Visit the [IPUMS USA website](#).
- 2. Register for an account or log in if you already have one.
- 3. Select the 2022 ACS dataset and choose relevant variables such as:
 - STATEICP: State codes for each respondent.
 - EDUCD: Educational attainment with detailed categories, including doctoral degrees.
- 4. Download the dataset in CSV format. The file may be compressed, so you need to unzip it using: `gunzip usa_00003.csv.gz`

3 Overview of the Ratio Estimators Approach

The ratio estimator approach is used when we want to estimate unknown totals in a population based on the known total in a sample and some measured characteristics.

We know the total number of respondents in California is 391,171. We can calculate the ratio of doctoral degree holders in California to the total respondents, then apply this ratio to estimate the total number of respondents in other states.

Ratio = Number of doctoral holders in California / 391171

We apply this ratio to the number of doctoral degree holders in other states to estimate their total respondent numbers.

4 Estimates and the Actual Number of Respondents

We computed the estimated total respondents based on the ratio estimator approach. Table 1 established by the tidyverse package (Wickham et al. 2019) shows the estimated number of respondents and the actual number of doctoral degree holders in each state.

Table 1: Estimated number of respondents and the actual number of doctoral degree holders in each state

State	Estimated Total Respondents	Doctoral Degree Holders	Difference
1	37042.708	600	36442.708
2	10186.745	165	10021.745
3	124340.024	2014	122326.024
4	15064.035	244	14820.035
5	10927.599	177	10750.599
6	8087.658	131	7956.658
11	9384.153	152	9232.153
12	88779.024	1438	87341.024
13	174656.370	2829	171827.370
14	100015.312	1620	98395.312
21	89952.043	1457	88495.043
22	38277.465	620	37657.465
23	61182.207	991	60191.207
24	74888.009	1213	73675.009
25	31671.516	513	31158.516
31	15928.365	258	15670.365
32	19817.849	321	19496.849
33	35314.049	572	34742.049

State	Estimated Total Respondents	Doctoral Degree Holders	Difference
34	38339.203	621	37718.203
35	9445.891	153	9292.891
36	3704.271	60	3644.271
37	4383.387	71	4312.387
40	94520.644	1531	92989.644
41	28399.410	460	27939.410
42	15496.200	251	15245.200
43	168606.061	2731	165875.061
44	89581.616	1451	88130.616
45	27782.031	450	27332.031
46	16237.054	263	15974.054
47	87729.481	1421	86308.481
48	39944.387	647	39297.387
49	198548.917	3216	195332.917
51	27658.556	448	27210.556
52	99274.458	1608	97666.458
53	17348.335	281	17067.335
54	51921.530	841	51080.530
56	9816.318	159	9657.318
61	55317.111	896	54421.111
62	63651.720	1031	62620.720
63	10804.123	175	10629.123
64	6976.377	113	6863.377
65	17410.073	282	17128.073
66	21608.247	350	21258.247
67	26423.799	428	25995.799
68	4445.125	72	4373.125
71	391171.000	6336	384835.000
72	39944.387	647	39297.387
73	73776.727	1195	72581.727
81	3148.630	51	3097.630
82	13211.899	214	12997.899
98	19200.470	311	18889.470

5 Explanation of Differences

There are several reasons why the estimated total number of respondents may differ from the actual numbers:

- State-level variations: The proportion of doctoral degree holders in California may not be representative of the proportion in other states. States with a higher or lower proportion of doctoral degree holders compared to California will have estimates that deviate from their actual totals.
- Population diversity: Each state has different demographic compositions. For instance, states with larger metropolitan areas or research universities may have a higher concentration of doctoral degree holders, skewing the estimates when compared to smaller or more rural states.
- Sampling variability: The IPUMS dataset is a sample of the population. Differences between the sample and the actual population (especially in smaller states) can cause the ratio estimator to be inaccurate.

References

- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Ruggles, Steven, Sarah Flood, Sophia Foster, Ronald Goeken, Jose Pacas, Megan Schouweiler, and Matthew Sobek. 2022. “IPUMS USA: Version 11.0.” Minneapolis, MN: IPUMS. <https://doi.org/10.18128/d010.v11.0>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Grolemond, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.