Estimating Total Respondents Using Ratio Estimator Approach*

Shamayla Durrin Denise Chang Krishna Kumar

November 21, 2024

Introduction

In this document, we use survey data from the 2022 American Community Survey (ACS) obtained through IPUMS (Ruggles et al. 2022). We apply the ratio estimator approach of Laplace to estimate the total number of respondents in each state based on the number of those with doctoral degrees. We compare these estimates to the actual number of respondents and discuss any differences.

How to Obtain the Data

- Go to https://www.ipums.org/
- Go to IPUMS USA
- Click 'Get Data'
- Click 'Select samples' and choose only '2022 ACS'
- Go to 'Household' -> 'Geographic' and click the + on 'STATEICP'
- Go to 'Person' -> 'Demographic' and click the + on 'SEX'
- Go to 'Person' -> 'Education' and click the + on 'EDUC'
- Click view cart
- Create data extract

*Code and data supporting this analysis are available at: https://github.com/DeniseChang9/UsingSurveyData

- Change data format to .csv
- Submit extract
- You will then get an email when the data is ready to download
- Download the data in your directory and read it using the code below.

Overview of the Ratio Estimators Approach

The Ratio Estimator Approach calculates the ratio of two means.

The ratio estimator is a method used to estimate the total size or characteristic of a population based on a sample (Borkowski n.d.). By calculating the ratio between a known variable and a target variable (for example, the ratio of doctoral degree holders to the total number of respondents in a sample), this ratio is then applied to a subgroup to estimate its total size. The method assumes that the ratio is relatively stable across different subgroups. For example, if 10% of California respondents have a doctoral degree, we can apply this ratio to other states to estimate their total respondents from the number with doctoral degrees.

This approach dates back to Quetelet and Laplace, who used similar methods to estimate populations based on partial data. A well-known variant is the capture-recapture method, which is commonly used in ecology. In this technique, a sample is captured, marked, and then released back into the population. Later, a second sample is captured, and the proportion of marked individuals is used to estimate the total population size. The ratio estimator is powerful when direct measurement of the whole population is impractical, allowing for accurate inference from smaller, representative samples.

Finding Total Respondents In State Using The Ratio Estimator

We will use the ratio estimator approach by first calculating the proportion of respondents with doctoral degrees to the total respondents in California. This ratio will be used as a benchmark for all other states. By applying this ratio to the number of doctoral degree holders in each state, we can estimate the total number of respondents for each state. This approach assumes that the proportion of doctoral degree holders is similar across states, allowing us to scale up from the known subset to estimate the full population.

- First we count the total number of respondents with a doctoral degree
- Then we calculate the total number of respondents in California
- Then we do total number of respondents with doctoral/total number of respondents in california
- Then we find the estimated total using our ratio: we do total number of doctoral divided by our california ratio

- Then we compare the estimated values with the actual values

Table 1: Comparison of Estimated and Actual Respondents by State

State ICP	Number of Doctoral Respondents	Estimated Total Respondents	Actual Total Respondents	Difference
1	600	37043	37369	-326
2	165	10187	14523	-4336
3	2014	124340	73077	51263
4	244	15064	14077	987
5	177	10928	10401	527
6	131	8088	6860	1228
11	152	9384	9641	-257
12	1438	88779	93166	-4387
13	2829	174656	203891	-29235
14	1620	100015	132605	-32590
21	1457	89952	128046	-38094
22	620	38277	69843	-31566
23	991	61182	101512	-40330
24	1213	74888	120666	-45778
25	513	31672	61967	-30295
31	258	15928	33586	-17658
32	321	19818	29940	-10122
33	572	35314	58984	-23670
34	621	38339	64551	-26212
35	153	9446	19989	-10543
36	60	3704	8107	-4403
37	71	4383	9296	-4913
40	1531	94521	88761	5760
41	460	28399	51580	-23181
42	251	15496	31288	-15792
43	2731	168606	217799	-49193
44	1451	89582	109349	-19767
45	450	27782	45040	-17258
46	263	16237	29796	-13559
47	1421	87729	109230	-21501
48	647	39944	54651	-14707
49	3216	198549	292919	-94370
51	448	27659	46605	-18946
52	1608	99274	62442	36832
53	281	17348	39445	-22097
54	841	51922	72374	-20452
56	159	9816	18135	-8319
61	896	55317	74153	-18836

State ICP	Number of Doctoral Respondents	Estimated Total Respondents	Actual Total Respondents	Difference
62	1031	63652	59841	3811
63	175	10804	19884	-9080
64	113	6976	11116	-4140
65	282	17410	30749	-13339
66	350	21608	20243	1365
67	428	26424	35537	-9113
68	72	4445	5962	-1517
71	6336	391171	391171	0
72	647	39944	43708	-3764
73	1195	73777	80818	-7041
81	51	3149	6972	-3823
82	214	13212	14995	-1783
98	311	19200	6718	12482

Why are the Actual Values and Estimated Values Different?

- The ratio is based on one state's data (California) so it must not uniformly apply to all states, as each state has different educational systems and each state has different demographics of people.
- The reason the estimates of total respondents differ from the actual numbers is due to our assumption that the ratio of respondents with doctoral degrees is constant across all states. In reality, this is not the case, as many states have varying levels of resources that impact higher education. Some states may lack access to higher education institutions, leading to a lower proportion of doctoral degree holders. Factors such as income disparities, state education policies, and local economic conditions can significantly influence the number of individuals pursuing advanced degrees. Consequently, states with fewer resources or less emphasis on higher education will have lower ratios of doctoral degree holders, causing our estimates to be off.

References

Borkowski, John. n.d. "Ratio and Regression Estimation." http://www.math.montana.edu/jobo/st446/documents/ho5a.pdf.

Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek. 2022. "IPUMS USA: Version 12.0 [American Community Survey 2022]." IPUMS, University of Minnesota. https://doi.org/10.18128/D010.V12.0.