

Estimation of Respondents in Each State Using IPUMS ACS 2022

Your Name

Invalid Date

1. Introduction

This document uses data from the 2022 American Community Survey (Steven Ruggles and Sobek (2022)) provided by IPUMS to estimate the total number of respondents in each U.S. state who have obtained a doctoral degree as their highest level of educational attainment. Python (Python Software Foundation (2023)) and its pandas (McKinney (2021)) package were used to manipulate and display the data. We apply the ratio estimator approach of Laplace (Cochran (1977)) to estimate the total number of respondents in each state based on a known total for California, which had 391,171 respondents across all levels of education. The results of this estimation are then compared to the actual number of respondents.

1.1 Data Source

The data used for this analysis comes from IPUMS (Steven Ruggles and Sobek (2022)), specifically the 2022 ACS dataset. To obtain this data, follow these steps:

1. Create an account at [IPUMS USA](#).
2. Use the data extraction tool and select the 2022 ACS dataset.
3. Include the relevant variables:
 - **STATEICP**: State Code (IPUMS variable for state).
 - **EDUC**: Educational attainment variable (filtering for doctoral degrees).
4. Once the dataset was processed it was downloaded to the **data** folder

2. Overview of the Ratio Estimators Approach

The ratio estimator approach is a common method in survey sampling that allows us to estimate totals based on known relationships between two variables. In this case, we use the ratio of respondents with a doctoral degree to the total number of respondents in each state. This ratio is then applied to the known total number of respondents in California (391,171) to estimate the total number of respondents in other states.

The ratio estimator is calculated as follows: $\hat{N}_i = \frac{\text{Doctoral Respondents}_i}{\text{Doctoral Respondents}_{CA}} \times \text{Total Respondents}_{CA}$

Where: - (\hat{N}_i) is the estimated total number of respondents in state (i) - ($\text{Doctoral Respondents}_i$) is the number of respondents with doctoral degrees in state (i) - $\text{Total Respondents}_{CA}$ is the known total number of respondents in California.

3. Analysis and Results

	state	n	doctoral_count	ratio
4	California	391171	6336	0.016198

Figure 1 This includes the data for california state exclusively showing approximately 1.6% of respondents have obtained a doctorate.

First let's just look at the state of California, we see here it has a ratio of approximately 1.6% of participants obtained a doctoral degree. Using this ratio as an estimator for the other states yields the results below.

	state	doctoral_count	n	ratio	estimated_count
0	Alabama	460	51580	0.008918	835.492840
1	Alaska	51	6972	0.007315	112.932456
2	Arizona	896	74153	0.012083	1201.130294
3	Arkansas	251	31288	0.008022	506.803024
4	California	6336	391171	0.016198	6336.187858
5	Colorado	1031	59841	0.017229	969.304518
6	Connecticut	600	37369	0.016056	605.303062
7	Delaware	152	9641	0.015766	156.164918
8	District of Columbia	311	6718	0.046294	108.818164
9	Florida	2731	217799	0.012539	3527.908202
10	Georgia	1451	109349	0.013269	1771.235102
11	Hawaii	214	14995	0.014271	242.889010
12	Idaho	175	19884	0.008801	322.081032
13	Illinois	1457	128046	0.011379	2074.089108

	state	doctoral_count	n	ratio	estimated_count
14	Indiana	620	69843	0.008877	1131.316914
15	Iowa	258	33586	0.007682	544.026028
16	Kansas	321	29940	0.010721	484.968120
17	Kentucky	448	46605	0.009613	754.907790
18	Louisiana	450	45040	0.009991	729.557920
19	Maine	165	14523	0.011361	235.243554
20	Maryland	1608	62442	0.025752	1011.435516
21	Massachusetts	2014	73077	0.027560	1183.701246
22	Michigan	991	101512	0.009762	1644.291376
23	Minnesota	572	58984	0.009698	955.422832
24	Mississippi	263	29796	0.008827	482.635608
25	Missouri	621	64551	0.009620	1045.597098
26	Montana	113	11116	0.010166	180.056968
27	Nebraska	153	19989	0.007654	323.781822
28	Nevada	282	30749	0.009171	498.072302
29	New Hampshire	244	14077	0.017333	228.019246
30	New Jersey	1438	93166	0.015435	1509.102868
31	New Mexico	350	20243	0.017290	327.896114
32	New York	2829	203891	0.013875	3302.626418
33	North Carolina	1421	109230	0.013009	1769.307540
34	North Dakota	60	8107	0.007401	131.317186
35	Ohio	1213	120666	0.010053	1954.547868
36	Oklahoma	281	39445	0.007124	638.930110
37	Oregon	647	43708	0.014803	707.982184
38	Pennsylvania	1620	132605	0.012217	2147.935790
39	Rhode Island	177	10401	0.017018	168.475398
40	South Carolina	647	54651	0.011839	885.236898
41	South Dakota	71	9296	0.007638	150.576608
42	Tennessee	841	72374	0.011620	1172.314052
43	Texas	3216	292919	0.010979	4744.701962
44	Utah	428	35537	0.012044	575.628326
45	Vermont	131	6860	0.019096	111.118280
46	Virginia	1531	88761	0.017249	1437.750678
47	Washington	1195	80818	0.014786	1309.089964
48	West Virginia	159	18135	0.008768	293.750730
49	Wisconsin	513	61967	0.008279	1003.741466
50	Wyoming	72	5962	0.012076	96.572476

4. Discussion of Results

The results show that the ratio estimators approach can provide a reasonable approximation of the total number of respondents in each state based on the number of doctoral degree holders. However, there are some differences between the estimated and actual totals.

These differences can be attributed to a variety of factors:

Sampling Variation: The ratio of doctoral respondents to total respondents may vary between states, leading to estimation errors. State-Specific Factors: States with a large population of highly educated individuals, such as Massachusetts or New York, may have higher ratios of doctoral degree holders, skewing the estimates. Limitations of the Ratio Estimator: The ratio estimator assumes that the relationship between doctoral degree holders and total respondents is constant across all states, which may not be the case in reality.

- Cochran, William G. 1977. *Sampling Techniques*. 3rd ed. New York: John Wiley & Sons.
- McKinney, Wes. 2021. *Pandas: Powerful Python Data Analysis Toolkit*. <https://pandas.pydata.org/>.
- Python Software Foundation. 2023. *Python: A Dynamic, Open Source Programming Language*. <https://www.python.org/>.
- Steven Ruggles, Miriam King, Sarah Flood, and Matthew Sobek. 2022. “IPUMS USA: Version 12.0 [dataset].” IPUMS, University of Minnesota. <https://usa.ipums.org/usa/>.