

IPUMS 2022 ACS Analysis

Dhruv Gupta, Justin Klip, Kevin Shen,

2024-10-03

Instructions to Obtain the Data

In order to acquire the data used for our IPUMS research study, we first navigated to (<https://www.ipums.org/>). Once on the webpage, click on the “Visit Site” button under IPUMS USA. Next, click on the white “Get Data” button under CREATE YOUR CUSTOM DATA SET.

Once on the webpage used for selecting samples and variables, click on the blue “Select Samples” button on the left. On the USA Samples tab, uncheck the “Default sample from each year” checkbox, and then reselect the checkbox corresponding to 2022. Once selected, click the “Submit Samples” button; this will take you to the variable selection webpage.

On this page, navigate to the “Household” dropdown menu under Select Harmonized Variables, and select “Geographic.” In the list of variables under Geographic Variables — Household, click the plus sign next to the “STATEICP” variable. Then, hover over the “Person” dropdown menu and select “Education.” In the list of variables under Education Variables — Person, click the plus sign next to “EDUC.”

In the Data Cart floating window, select “View Cart.” On the Data Cart webpage, select the blue “Create Data Extract” button. On the Extract Request webpage, to the right of Data Format, click “Change” and switch the format to .csv. Finally, press the “Submit Extract” button at the bottom of the page.

On the Download or Revise Extracts webpage, find the extract you requested, and in the Download Data column, click “Download .csv.” This will give you the data necessary to run the code for our analysis.

Overview of Ratio Estimator Approach

We try to use the ratio estimators’ approach to find total respondents for each state using information we have available. We look at the number of total doctoral degree holders in California and compare it to the total number of respondents in California. We find there are

about 61.74 times as many people compared to doctoral degrees in California. So, we simply multiply the number of doctoral degrees holders in each state by this ratio (61.74) to get our estimate of total respondents per state. The assumption is that other states would have similar ratios of PhD's to total respondents. This approach is necessary if we only have doctoral data for all the states, but not total respondent data.

Comparison of Estimated Totals and Actual Totals

Below is a table with relevant

Table 1: Sample Comparison of Estimated and Actual Total Respondents for Selected States

STATEICP	doctoral_count	Estimated_Total_Respondents	Actual_Total_Respondents
3	2014	124340.024	73077
13	2829	174656.370	203891
41	460	28399.410	51580
42	251	15496.200	31288
68	72	4445.125	5962
71	6336	391171.000	391171

Explanation as to Differing Results:

Some states may get differing values due to differing state characteristics. We based our ratio on California, which has a large number of educational institutes that award PhD's. California's large academic and advanced labor market may also draw in a lot of PhD's in comparison to a state like Wyoming, which has less incentive for PhD's to live there. This differing state level trends may explain why we get close estimates for some states, but not for others. In Wyoming (Code 68) we predict there to be a much smaller number of people (4445.125) than there actually are (5962) and this is because of this discrepancy in assuming there are much more PhD's in Wyoming then there are, so we multiply by too high of a number. Other states also underestimate suggesting the same issue as Wyoming. Massachusetts (Code 03), has an overestimate of the number of people, this is likely because they have a higher ratio of PhD graduates in their total population. This makes sense given that they have some of the best research institutions in the world that attract PhD's, so in Massachusetts our ratio is an underestimate.

Appendix

Table 2: Complete Comparison of Estimated and Actual Total Respondents by State

STATEICP	doctoral_count	Estimated_Total_Respondents	Actual_Total_Respondents
1	600	37042.708	37369
2	165	10186.745	14523
3	2014	124340.024	73077
4	244	15064.035	14077
5	177	10927.599	10401
6	131	8087.658	6860
11	152	9384.153	9641
12	1438	88779.024	93166
13	2829	174656.370	203891
14	1620	100015.312	132605
21	1457	89952.043	128046
22	620	38277.465	69843
23	991	61182.207	101512
24	1213	74888.009	120666
25	513	31671.516	61967
31	258	15928.365	33586
32	321	19817.849	29940
33	572	35314.049	58984
34	621	38339.203	64551
35	153	9445.891	19989
36	60	3704.271	8107
37	71	4383.387	9296
40	1531	94520.644	88761
41	460	28399.410	51580
42	251	15496.200	31288
43	2731	168606.061	217799
44	1451	89581.616	109349
45	450	27782.031	45040
46	263	16237.054	29796
47	1421	87729.481	109230
48	647	39944.387	54651
49	3216	198548.917	292919
51	448	27658.556	46605
52	1608	99274.458	62442

53	281	17348.335	39445
54	841	51921.530	72374
56	159	9816.318	18135
61	896	55317.111	74153
62	1031	63651.720	59841
63	175	10804.123	19884
64	113	6976.377	11116
65	282	17410.073	30749
66	350	21608.247	20243
67	428	26423.799	35537
68	72	4445.125	5962
71	6336	391171.000	391171
72	647	39944.387	43708
73	1195	73776.727	80818
81	51	3148.630	6972
82	214	13211.899	14995
98	311	19200.470	6718
