

IPUMS 2022 ACS Data Analysis Using Ratio Estimation

Charlie Zhang

October 3, 2024

1 Instructions on How to Obtain the Data

To obtain the IPUMS 2022 ACS data: 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: `bash gunzip usa_00003.csv.gz` # 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.

2 Estimates and the Actual Number of Respondents

We computed the estimated total respondents based on the ratio estimator approach. Below is the table showing the estimated number of respondents and the actual number of doctoral degree holders in each state.

| STATEICP | estimate_total | count_doctoral | 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 |
| 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 |

| STATEICP | estimate_total | count_doctoral | difference |
|----------|----------------|----------------|------------|
| 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 |

3 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.