

Estimation of Doctors in Each State

Yanzun Jiang

Instructions on obtaining the data

1. Go to the **IPUMS website** and click on “VISIT SITE” under the IPUMS USA section.
2. Click on “**Get Data**”.
3. Click “**Select Samples**”, uncheck “**Default sample from each year**”, and check only “**ACS**” for the year 2022.
4. Click “**Submit Sample Selections**”.
5. Under **HOUSEHOLD**, hover over “**GEOGRAPHIC**” and add “**STATEICP**” to your cart.
6. Under **PERSON**, hover over “**DEMOGRAPHIC**” and add “**SEX**” to your cart.
7. Under **PERSON**, hover over “**EDUCATION**” and add “**EDUC**” to your cart.
8. In your **Data Cart**, you should now see 4 variables and 1 sample. Click “**View Cart**”.
9. Click “**Create Data Extract**”.
10. Change the **Data Format** from .dat to .csv and click “**Apply Selections**”.
11. Click “**Submit Extract**”.
12. Wait for the status to become **COMPLETED**, then click “**Download .csv**” under the “Download Data” column.
13. Finally, unzip the downloaded file to access your data.

Overview of the ratio estimators approach

1. Let:

- R_{CA} be the ratio of the number of people with doctoral degrees to total respondents in CA.

- $N_{doctoral_degrees,state}$ be the number people with of doctoral degrees in any other state.

- $N_{respondents,state}$ be the estimated number of respondents in other states.

2. Application to Other States: For any other state, the estimated number of total respondents can be expressed as:

$$N_{respondents,state} = \frac{N_{doctoral_degrees,state}}{R_{CA}}$$

This uses the ratio observed in California to estimate the total number of respondents based on the number of doctoral degrees in other states.

Summary table of estimated and actual respondents

Table 1: Estimated and Actual Respondents

state_id	estimated	actual	diff
NA	37043	37369	326
NA	10187	14523	4336
NA	124340	73077	-51263
NA	15064	14077	-987
NA	10928	10401	-527
NA	8088	6860	-1228
NA	9384	9641	257
NA	88779	93166	4387
NA	174656	203891	29235
NA	100015	132605	32590
NA	89952	128046	38094
NA	38277	69843	31566
NA	61182	101512	40330
NA	74888	120666	45778
NA	31672	61967	30295
NA	15928	33586	17658
NA	19818	29940	10122
NA	35314	58984	23670
NA	38339	64551	26212

Table 1: Estimated and Actual Respondents

state_id	estimated	actual	diff
NA	9446	19989	10543
NA	3704	8107	4403
NA	4383	9296	4913
NA	94521	88761	-5760
NA	28399	51580	23181
NA	15496	31288	15792
NA	168606	217799	49193
NA	89582	109349	19767
NA	27782	45040	17258
NA	16237	29796	13559
NA	87729	109230	21501
NA	39944	54651	14707
NA	198549	292919	94370
NA	27659	46605	18946
NA	99274	62442	-36832
NA	17348	39445	22097
NA	51922	72374	20452
NA	9816	18135	8319
NA	55317	74153	18836
NA	63652	59841	-3811
NA	10804	19884	9080
NA	6976	11116	4140
NA	17410	30749	13339
NA	21608	20243	-1365
NA	26424	35537	9113
NA	4445	5962	1517
NA	NA	391171	NA
NA	39944	43708	3764
NA	73777	80818	7041
NA	3149	6972	3823
NA	13212	14995	1783
NA	19200	6718	-12482

Explanation of the differences

If we estimate the total number of respondents using the ratio estimator approach with Laplace's method, we are assuming that the proportion of people with doctoral degrees is

the same across all states. However, this is not necessarily the case. California, with its many large universities and research institutions, likely has a higher concentration of individuals with doctoral degrees. As a result, the doctor-to-respondent ratio in California would be higher than in other states, leading to an underestimate of the total number of respondents in those states.