Estimation of Doctors in Each State

Group 51

October 3, 2024

Table of contents

1 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.

2 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.

3 Summary Table of Estimated and Actual Respondents

Table 1: Estimated and Actual Respondents

state_id	estimated	actual	diff
Connecticut	37043	37369	326
Maine	10187	14523	4336
Massachusetts	124340	73077	-51263
New Hampshire	15064	14077	-987
Rhode Island	10928	10401	-527
Vermont	8088	6860	-1228
Delaware	9384	9641	257
New Jersey	88779	93166	4387
New York	174656	203891	29235
Pennsylvania	100015	132605	32590
Illinois	89952	128046	38094
Indiana	38277	69843	31566
Michigan	61182	101512	40330
Ohio	74888	120666	45778
Wisconsin	31672	61967	30295
Iowa	15928	33586	17658
Kansas	19818	29940	10122
Minnesota	35314	58984	23670
Missouri	38339	64551	26212
Nebraska	9446	19989	10543
North Dakota	3704	8107	4403
South Dakota	4383	9296	4913
Virginia	94521	88761	-5760
Alabama	28399	51580	23181
Arkansas	15496	31288	15792
Florida	168606	217799	49193
Georgia	89582	109349	19767
Louisiana	27782	45040	17258
Mississippi	16237	29796	13559
North Carolina	87729	109230	21501
South Carolina	39944	54651	14707
Texas	198549	292919	94370
Kentucky	27659	46605	18946
Maryland	99274	62442	-36832
Oklahoma	17348	39445	22097
Tennessee	51922	72374	20452

Table 1: Estimated and Actual Respondents

state_id	estimated	actual	diff
West Virginia	9816	18135	8319
Arizona	55317	74153	18836
Colorado	63652	59841	-3811
Idaho	10804	19884	9080
Montana	6976	11116	4140
Nevada	17410	30749	13339
New Mexico	21608	20243	-1365
Utah	26424	35537	9113
Wyoming	4445	5962	1517
California	391171	391171	0
Oregon	39944	43708	3764
Washington	73777	80818	7041
Alaska	3149	6972	3823
Hawaii	13212	14995	1783
District of Columbia	19200	6718	-12482

4 Explanation of 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.