Reflection 4

Group 8

October 3, 2024

1 Instruction on data gathering

1. Visit the IPUMS USA homepage https://usa.ipums.org/usa/, to access U.S. Census and American Community Survey microdata from 1850 to the present.
2. Go to Online Tool for Analysis – Analyze Data Online.
3. We are interested in accessing the 2022 ACS to examine "How many respondents were there in each state that had a doctoral degree as their highest educational attainment". In the top right corner, go to Select Data.
4. Under Select Source Variables, select the Sample to be from 2022 ACS. We selected the variables:

- US2022A_YEAR
- US2022A_SAMPLE
- US2022A_SERIAL
- US2022A_CBSERIAL
- US2022A_HHWT
- US2022A_CLUSTER
- US2022A_STRATA
- US2022A_GQ
- US2022A_STATEICP
- US2022A_PERNUM
- US2022A_PERWT
- US2022A_SEX
- US2022A_EDUC (general)
- US2022A_EDUCD (detailed)

5. View your cart and click Create Data Extract.
6. Check over your Samples and Variables by clicking the (show) option to ensure that you have enough variables for the desired exploration.

7. Under 'DATA FORMAT', change to .csv (Comma delimited) and apply selections.
8. Now we click Submit Extract. Under 'DOWNLOAD DATA', we should have the file ready for download. If an error has occurred, visit the 'REVISE EXTRACT' column and head back to step 6.
9. Finally, we can extract the zip file and begin in R.

2 Overview of the ratio estimators approach

In this reflection, we used Laplace's ratio estimator approach to estimate the population of each state in the U.S. Since the total number of respondents for California is provided, we used California as the "base." First, we estimated the ratio by dividing the number of people with doctoral degrees in each state by the number of people with doctoral degrees in California. Then, we multiplied these ratios by the total number of respondents in California to estimate the population of each state.

3 Estimates and Actual Number of Respondents

Table 1: Estimates and Actual Number of Respondents

State Code	Actual Population	Estimated Population
1	37369	37042.708
2	14523	10186.745
3	73077	124340.024
4	14077	15064.035
5	10401	10927.599
6	6860	8087.658
11	9641	9384.153
12	93166	88779.024
13	203891	174656.370
14	132605	100015.312
21	128046	89952.043
22	69843	38277.465
23	101512	61182.207
24	120666	74888.009
25	61967	31671.516
31	33586	15928.365
32	29940	19817.849
33	58984	35314.049

Table 1: Estimates and Actual Number of Respondents

State Code	Actual Population	Estimated Population
34	64551	38339.203
35	19989	9445.891
36	8107	3704.271
37	9296	4383.387
40	88761	94520.644
41	51580	28399.410
42	31288	15496.200
43	217799	168606.061
44	109349	89581.616
45	45040	27782.031
46	29796	16237.054
47	109230	87729.481
48	54651	39944.387
49	292919	198548.917
51	46605	27658.556
52	62442	99274.458
53	39445	17348.335
54	72374	51921.529
56	18135	9816.318
61	74153	55317.111
62	59841	63651.721
63	19884	10804.123
64	11116	6976.377
65	30749	17410.073
66	20243	21608.247
67	35537	26423.799
68	5962	4445.125
71	391171	391171.000
72	43708	39944.387
73	80818	73776.727
81	6972	3148.630
82	14995	13211.899
98	6718	19200.470

4 Why they are different?

1. We are assuming that the ratio between the number of respondents with doctoral degrees in California and the number of respondents in California is the same across all states, which is not true since local factors such as population demographics, educational opportunities, or economic conditions are different across states.
2. Sample size is different for each state, hence the ratio is different
3. Errors in data collection or reporting may lead to differences between estimated and actual count.