reflection of Laplace for phd in USA

Making use of the codebook, how many respondents were there in each state (STATEICP) that had a doctoral degree as their highest educational attainment (EDUC)? (Hint: Make this a column in a tibble.)

Table 1: Count of Doctoral Degree Holders by State

STATEICP	doctoral_count
1	600
2	165
3	2014
4	244
5	177
6	131
11	152
12	1438
13	2829
14	1620
21	1457
22	620
23	991
24	1213
25	513
31	258
32	321
33	572
34	621
35	153
36	60
37	71
40	1531

41	460
42	251
43	2731
44	1451
45	450
46	263
47	1421
48	647
49	3216
51	448
52	1608
53	281
54	841
56	159
61	896
62	1031
63	175
64	113
65	282
66	350
67	428
68	72
71	6336
72	647
73	1195
81	51
82	214
98	311

Instructions for obtaining the data.

To obtain data from IPUMS USA, we first navigated to the IPUMS website and selected "IPUMS USA." Next, we clicked on "Get Data" and chose "2022 ACS" under the "SELECT SAMPLE" section. For state-level data, we selected "HOUSEHOLD," then "GEOGRAPHIC," and added "STATEICP" to our cart. For individual-level data, we selected "PERSON," then navigated to "DEMOGRAPHIC" and "EDUCATION" where we added "SEX" and "EDUC" to our cart respectively. Afterward, we clicked "VIEW CART" followed by "CREATE DATA EXTRACT." We changed the "DATA FORMAT" to ".csv" and clicked "SUBMIT EXTRACT."

After creating an account and verifying it, received an email notification when the extract was ready. Finally, we downloaded and saved the file locally (e.g., "usa_00001.csv.gz") for use in R.

A concise summary of the ratio estimators method.

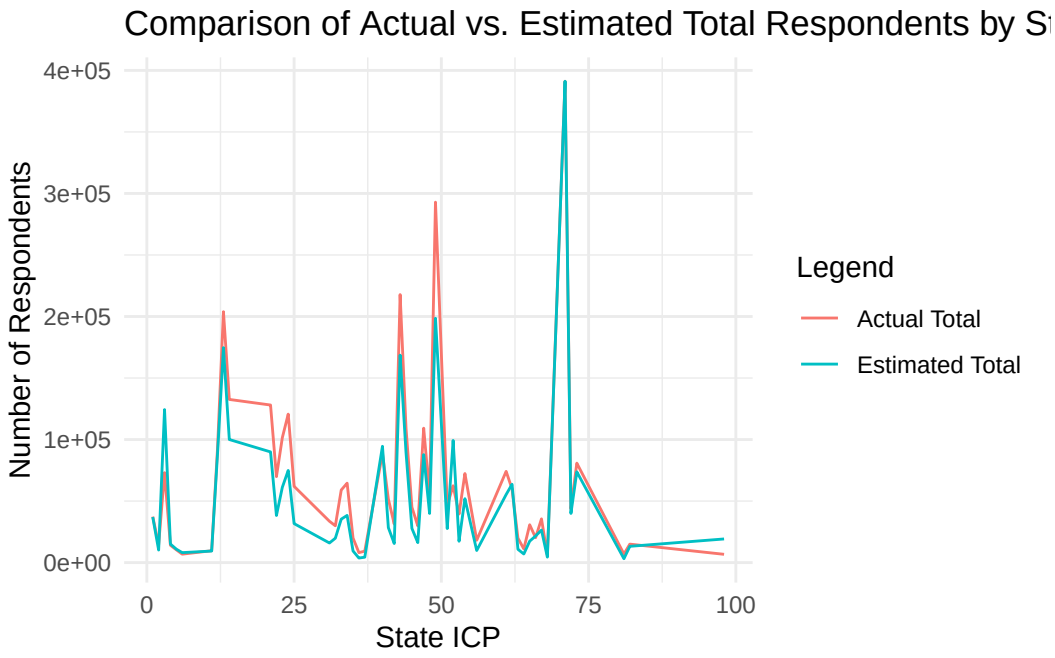
Ratio estimation is a statistical method that estimates population totals or means by using known ratios from a sample. By calculating the ratio of a characteristic (e.g., individuals with a PhD) to the total population in a known subset (e.g., California), researchers can apply this ratio to other subsets to infer broader population totals. This approach is particularly useful when direct population measurement is impractical, enhancing estimation accuracy, especially when there's a strong correlation between the variable of interest and an auxiliary variable.

The actual number of respondents and your estimations.

Table 2: Compare of actual and estimated Holders by State

STATEICP	actual_total	estimated_total
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
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	51921.530
56	18135	9816.318
61	74153	55317.111
62	59841	63651.720
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



A brief reasoning for what you think that they differ.

The estimated total number of respondents in each state using the ratio estimators approach may differ from the actual number of respondents for several reasons:

- Ratio estimation methods typically assume that relationships observed in one subset of the data are representative of other subsets. However, this assumption often does not hold in real world. States may have very different demographics, economic conditions, and educational policies, making it inaccurate to generalize patterns of educational attainment in one state to other states.
- Educational attainment across U.S. states varies significantly due to differences in regional policies, cultural factors, and access to economic opportunities. Some states invest more in higher education, offering greater resources and opportunities, while others may prioritize vocational training or immediate entry into the workforce due to economic or cultural reasons. California, with its large population and abundant resources, represents a unique case where its educational patterns may not be generalizable to other states with differing economic conditions and cultural contexts. Thus, using California's ratio to estimate national trends could lead to inaccurate conclusions.
- The ratio estimation method relies on the assumption that there is a consistent relationship between the characteristics of doctoral degrees and the total population in each state. If this relationship is inconsistent, the estimates derived from this method will be biased. For example, states with different levels of public education funding may

have different relationships between population and educational attainment, resulting in biased estimates.

These reasons indicate that the assumption of homogeneity used in ratio estimators often leads to differences when applied to diverse populations such as different states in the US.

Reference