Analysis of Doctoral Degrees Cross the United States

Haowei Fan Tianning He Julia Lee Shuangyuan Yang
Tiffany Kim Max Li

Introduction

The data was obtained from IPUMS USA (IPUMS 2024), and the datatset used within this analysis contains information from the 2022 American Community Survey (ACS). Through the IPUM website, the 2022 ACS along with the variables, education and sex, were selected and downloaded as a csv file. We are interested in how many respondents in California (STATEICP) have a doctorate as their highest level of education (EDUC). We know that there are 391,171 respondents of all educational levels in California (STATEICP). We searched the IPUMS website for the total population and education of all states in the United States, and screened out the total population of California and the number of people with a doctorate. We hope to estimate the total number of respondents in each state using the Laplace ratio estimation method. All analyses to understand the number of residents who have earned a doctoral degree and estimate the number of respondents for every state were conducted using the statistical programming language R (R Core Team 2023).

Overview of the ratio estimators approach

The ratio estimator method is a statistical estimating tool for totals or averages by using known relationships between variables from a sample. It calculates the proportion of a specific characteristic such as doctoral degree holders to the total population in a known group like California. This proportion is then extended to estimate figures for other groups, based on the assumption that similar relationships exist throughout the population. This method is especially useful when the exact population size is unknown but can be inferred through sample proportions.

Comparison of the actual data and our estimated data.

Shown below in Table 1.

Table 1: Comparison

State	Total respondents	Estimated total respondents	Errors
41	51580	28399	23181
81	6972	3149	3823
61	74153	55317	18836
42	31288	15496	15792
71	391171	391171	0
62	59841	63652	-3811
1	37369	37043	326
11	9641	9384	257
98	6718	19200	-12482
43	217799	168606	49193
44	109349	89582	19767
82	14995	13212	1783
63	19884	10804	9080
21	128046	89952	38094
22	69843	38277	31566
31	33586	15928	17658
32	29940	19818	10122
51	46605	27659	18946
45	45040	27782	17258
2	14523	10187	4336
52	62442	99274	-36832
3	73077	124340	-51263
23	101512	61182	40330
33	58984	35314	23670
46	29796	16237	13559
34	64551	38339	26212
64	11116	6976	4140
35	19989	9446	10543
65	30749	17410	13339
4	14077	15064	-987
12	93166	88779	4387
66	20243	21608	-1365
13	203891	174656	29235
47	109230	87729	21501
36	8107	3704	4403

Table 1: Comparison

State	Total respondents	Estimated total respondents	Errors
24	120666	74888	45778
53	39445	17348	22097
72	43708	39944	3764
14	132605	100015	32590
5	10401	10928	-527
48	54651	39944	14707
37	9296	4383	4913
54	72374	51922	20452
49	292919	198549	94370
67	35537	26424	9113
6	6860	8088	-1228
40	88761	94521	-5760
73	80818	73777	7041
56	18135	9816	8319
25	61967	31672	30295
68	5962	4445	1517

Interpretation of Our Comparison

1. The education system varies in each state, meaning that the number of doctoral degrees in California that was used within the ratio we found would not be representative of all states, and therefore, result in the discrepancy seen between our estimates and the actual data.
2. Also, the total number of respondents in California was given to us for this analysis, and as this number could be different in actuality, the estimate we obtained would not reflect the true respondent count within each state.

References

IPUMS. 2024. *Codebook for an IPUMS USA Data Extract.* Vienna, Austria: University of Minnesota. https://ipums.org.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.