Analysis of Doctoral Degrees Cross the United States*

Kevin Roe Cher Ning-Li Fangning Zhang Jierui Zhan
Jingchuan Xu

November 21, 2024

Table of contents

1 How to Obtain the Data

The data was obtained from IPUMS USA (IPUMS 2024). After creating the an IPUMS USA account, you first go to the Get Data button on the home page. Then you want to Select Samples and only select ACS 2022 option and deselect all other default options and you submit your sample selection. Under Select Harmonized Variables and Geographic, select STATEICP. Under Demographic choose SEX, EDUC, EDUCD. You then go under View Cart and create a data extract and make sure to create the data extract as a csv. Once you're able to, download and unzip the data.

*All data, R code, and other files are available in the following GitHub repository: https://github.com/Kanghyunroe/us-postdocs

Unzip the data and look for a file called "usa_00002.csv". Load the file in the quarto document and work off the document.

All analyses of this resesarch were conducted using the statistical programming language R (R Core Team 2023).

2 Overview of Ratio Estimators Approach

The ratio estimators approach first assumes the number of respondents in California then finds the percentage of respondents with a doctorate. Using this ratio, assuming that all states ahve the same percentage of doctorates, we divide the number of doctorates by the ratio to find the estimated number of respondents for each state.

3 Estimates and Actual Number of Respondents

Please find results under Section 5

4 Explanation for Difference

The reason why there is a difference is because the Ratio Estimators Approach assumes that the percentage of doctorates across all 50 states equals the percentage of doctorates in California. California, a highly educated state, is not a good representation of the entire population in terms of the proportion of doctorates. With many research universities and start-ups the proportion of doctorates in California does not represent the entire population. Therefore, because we use the California proportion for all states, it is not a good representation and there will be discrepancies.

5 Code Output

Table 1: Number of Doctoral Degree Holders by State

State Code	Number of Doctorates
1	600
2	165
3	2014
4	244
5	177

State Code	Number of Doctorates
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

State Code	Number of Doctorates
72	647
73	1195
81	51
82	214
98	311

Table 2: Estimated Total Number of Respondents by State

State Code	Number of Doctorates	Estimated Total
1	600	37043
2	165	10187
3	2014	124340
4	244	15064
5	177	10928
6	131	8088
11	152	9384
12	1438	88779
13	2829	174656
14	1620	100015
21	1457	89952
22	620	38277
23	991	61182
24	1213	74888
25	513	31672
31	258	15928
32	321	19818
33	572	35314
34	621	38339
35	153	9446
36	60	3704
37	71	4383
40	1531	94521
41	460	28399
42	251	15496
43	2731	168606
44	1451	89582
45	450	27782
46	263	16237
47	1421	87729
48	647	39944

State Code	Number of Doctorates	Estimated Total
49	3216	198549
51	448	27659
52	1608	99274
53	281	17348
54	841	51922
56	159	9816
61	896	55317
62	1031	63652
63	175	10804
64	113	6976
65	282	17410
66	350	21608
67	428	26424
68	72	4445
71	6336	391171
72	647	39944
73	1195	73777
81	51	3149
82	214	13212
98	311	19200

Table 3: Comparison Between Estimated and Actual Respondents by State

State Code	Number of Doctorates	Estimated Respondents	Actual Respondents	Difference
1	600	37043	37369	326
2	165	10187	14523	4336
3	2014	124340	73077	-51263
4	244	15064	14077	-987
5	177	10928	10401	-527
6	131	8088	6860	-1228
11	152	9384	9641	257
12	1438	88779	93166	4387
13	2829	174656	203891	29235
14	1620	100015	132605	32590
21	1457	89952	128046	38094
22	620	38277	69843	31566
23	991	61182	101512	40330
24	1213	74888	120666	45778
25	513	31672	61967	30295

State Code	Number of Doctorates	Estimated Respondents	Actual Respondents	Difference
31	258	15928	33586	17658
32	321	19818	29940	10122
33	572	35314	58984	23670
34	621	38339	64551	26212
35	153	9446	19989	10543
36	60	3704	8107	4403
37	71	4383	9296	4913
40	1531	94521	88761	-5760
41	460	28399	51580	23181
42	251	15496	31288	15792
43	2731	168606	217799	49193
44	1451	89582	109349	19767
45	450	27782	45040	17258
46	263	16237	29796	13559
47	1421	87729	109230	21501
48	647	39944	54651	14707
49	3216	198549	292919	94370
51	448	27659	46605	18946
52	1608	99274	62442	-36832
53	281	17348	39445	22097
54	841	51922	72374	20452
56	159	9816	18135	8319
61	896	55317	74153	18836
62	1031	63652	59841	-3811
63	175	10804	19884	9080
64	113	6976	11116	4140
65	282	17410	30749	13339
66	350	21608	20243	-1365
67	428	26424	35537	9113
68	72	4445	5962	1517
71	6336	391171	391171	0
72	647	39944	43708	3764
73	1195	73777	80818	7041
81	51	3149	6972	3823
82	214	13212	14995	1783
98	311	19200	6718	-12482

reference

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/.