

# Week 5 Reflection\*

Yunkai Gu, Anqi Xu, Yitong Wang

October 3, 2024

## Table of contents

<b>1 Data obtaining</b>	<b>1</b>
<b>2 Overview of the ratio estimators approach</b>	<b>3</b>
<b>3 Estimates and the actual number of respondents</b>	<b>3</b>
<b>4 Reason of Difference</b>	<b>6</b>

## 1 Data obtaining

We gather the data from IPUMS USA site, firstly we select “IPUMS USA” on the IPUMS, then clicked “Get Data”, then click “SELECT SAMPLE” and only select “2022 ACS”. We choose state level data by selecting “HOUSEHOLD”, then choose “GEOGRAPHIC” and add “STATEICP” to cart. For individual level data, we directly search “EDUC” and add it to the cart. After that, we clicked “VIEW CART”, then click “CREATE DATA EXTRACT”. We modify the it to csv form. We clicked “SUBMIT EXTRACT” and download it.

Table 1 shows the number of respondents that had a doctoral degree as their highest educational attainment (EDUC) in each state (STATEICP).

---

\*Code and data are available at: <https://github.com/Anjojoo/Reflection-4>.

Table 1: Respondents with a doctoral degree as their highest educational attainment in each 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

Table 1: Respondents with a doctoral degree as their highest educational attainment in each state

STATEICP	doctoral_count
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

## 2 Overview of the ratio estimators approach

The ratio estimator is a method used to improve the accuracy of estimates for a population parameter when there is an auxiliary variable related to the variable of interest. In this case, the objective is to estimate the total number of respondents in each state in the 2022 ACS dataset, given the known number of respondents with doctoral degrees in each state and the California ratio.

With the given total number of respondents in California across all education levels and the number of respondents in California who have a doctoral degree which is available in the data, we can calculate the ratio by the following:

$$Ratio = \frac{\text{Total number of respondents}}{\text{Number of doctoral respondents}}$$

Once the ratio is known for California, it is assumed that this ratio is similar across other states. This is the core assumption of the ratio estimator: that the proportion of doctoral degree holders to total respondents is similar across states.

For each state, the estimated total number of respondents is calculated by applying the ratio derived from California:

$$\text{Estimated Total Respondents in State} = \frac{\text{Number of doctoral respondents in state}}{\text{Ratio}}$$

### 3 Estimates and the actual number of respondents

Table 2 shows the number of estimated total respondents in each state by estimators approach of Laplace.

Table 2: Number of Estimated Total Respondents in Each State

STATEICP	estimated_total_respondents
1	37042.708
2	10186.745
3	124340.024
4	15064.035
5	10927.599
6	8087.658
11	9384.153
12	88779.024
13	174656.370
14	100015.312
21	89952.043
22	38277.465
23	61182.207
24	74888.009
25	31671.516
31	15928.365
32	19817.849
33	35314.049
34	38339.203
35	9445.891
36	3704.271
37	4383.387
40	94520.644
41	28399.410
42	15496.200
43	168606.061
44	89581.616
45	27782.031
46	16237.054
47	87729.481
48	39944.387
49	198548.917
51	27658.556
52	99274.458

Table 2: Number of Estimated Total Respondents in Each State

STATEICP	estimated_total_respondents
53	17348.335
54	51921.530
56	9816.318
61	55317.111
62	63651.720
63	10804.123
64	6976.377
65	17410.073
66	21608.247
67	26423.799
68	4445.125
71	391171.000
72	39944.387
73	73776.727
81	3148.630
82	13211.899
98	19200.470

Table 3 shows the actual respondent and the difference between estimation and the actual number of respondents in each state.

Table 3: Number of Actual Total Respondents and the Difference in Each State

STATEICP	estimated_total_respondents	total_count_respondents	difference
1	37042.708	37369	326.2917
2	10186.745	14523	4336.2552
3	124340.024	73077	-51263.0243
4	15064.035	14077	-987.0347
5	10927.599	10401	-526.5990
6	8087.658	6860	-1227.6580
11	9384.153	9641	256.8472
12	88779.024	93166	4386.9757
13	174656.370	203891	29234.6302
14	100015.312	132605	32589.6875
21	89952.043	128046	38093.9566
22	38277.465	69843	31565.5347
23	61182.207	101512	40329.7934
24	74888.009	120666	45777.9913

Table 3: Number of Actual Total Respondents and the Difference in Each State

STATEICP	estimated_total_respondents	total_count_respondents	difference
25	31671.516	61967	30295.4844
31	15928.365	33586	17657.6354
32	19817.849	29940	10122.1510
33	35314.049	58984	23669.9514
34	38339.203	64551	26211.7969
35	9445.891	19989	10543.1094
36	3704.271	8107	4402.7292
37	4383.387	9296	4912.6128
40	94520.644	88761	-5759.6441
41	28399.410	51580	23180.5903
42	15496.200	31288	15791.8003
43	168606.061	217799	49192.9392
44	89581.616	109349	19767.3837
45	27782.031	45040	17257.9688
46	16237.054	29796	13558.9462
47	87729.481	109230	21500.5191
48	39944.387	54651	14706.6128
49	198548.917	292919	94370.0833
51	27658.556	46605	18946.4444
52	99274.458	62442	-36832.4583
53	17348.335	39445	22096.6649
54	51921.530	72374	20452.4705
56	9816.318	18135	8318.6823
61	55317.111	74153	18835.8889
62	63651.720	59841	-3810.7205
63	10804.123	19884	9079.8767
64	6976.377	11116	4139.6233
65	17410.073	30749	13338.9271
66	21608.247	20243	-1365.2465
67	26423.799	35537	9113.2014
68	4445.125	5962	1516.8750
71	391171.000	391171	0.0000
72	39944.387	43708	3763.6128
73	73776.727	80818	7041.2726
81	3148.630	6972	3823.3698
82	13211.899	14995	1783.1007
98	19200.470	6718	-12482.4705

## 4 Reason of Difference

As the primary assumption is that the ratio of doctoral degree holders to total respondents in California is similar to that in other states. If this is not the case, the ratio estimator can produce biased estimates. Differences in educational attainment by state due to factors like demographics, local policies, and economy may cause the actual ratios to differ, leading to discrepancies between the estimates and actual values.