Reflection 4 Exercise*

Group 52

October 3, 2024

1 Instructions on how to obtain the data:

1. Go to https://usa.ipums.org/usa/
2. Create an account or log in
3. Select the 2022 ACS sample
4. Choose the following variables: STATEICP, EDUC, SEX
5. Submit the extract request
6. Download the data and save it as "usa_00001.csv" in a "data" folder in your project directory

2 Brief Overview of the Ratio Estimators Approach

The ratio estimators approach, also known as the Laplace ratio estimator, is a statistical method used to estimate population parameters when only partial information is available. In this case, we're using it to estimate the total number of respondents in each state based on the known number of respondents with doctoral degrees.

The basic idea behind this approach is to use a known ratio from one population (in this case, California) and apply it to other populations to estimate their total size. The steps involved are:

1. Calculate the ratio of doctoral degree holders to total respondents in California. (Assume this ratio is constant across all states)
2. For each state, divide the number of doctoral degree holders by this ratio to estimate the total number of respondents.

This method relies on the assumption that the proportion of doctoral degree holders is relatively consistent across states, which may not always be true in practice.

*Code and data are available at: https://github.com/stevenli-uoft/sta304_reflection4

3 Estimates and Actual Number of Respondents

Table 1 presents the total doctoral count, total respondents, and estimated respondent count for every state.

Table 1: State Doctoral and Respondant Counts, and Estimates

STATEICP	Actual Doctoral Count	Total Respondent	Estimated Respondent Count	Difference	% Difference
71	6336	391171	391171.000	0.0000	0.0000000
49	3216	292919	198548.917	94370.0833	32.2171260
13	2829	203891	174656.370	29234.6302	14.3383623
43	2731	217799	168606.061	49192.9392	22.5863935
3	2014	73077	124340.024	-51263.0243	-70.1493278
14	1620	132605	100015.312	32589.6875	24.5765148
52	1608	62442	99274.458	-36832.4583	-58.9866730
40	1531	88761	94520.644	-5759.6441	-6.4889356
21	1457	128046	89952.043	38093.9566	29.7502121
44	1451	109349	89581.616	19767.3837	18.0773337
12	1438	93166	88779.024	4386.9757	4.7087733
47	1421	109230	87729.481	21500.5191	19.6837124
24	1213	120666	74888.009	45777.9913	37.9377715
73	1195	80818	73776.727	7041.2726	8.7125053
62	1031	59841	63651.720	-3810.7205	-6.3680762
23	991	101512	61182.207	40329.7934	39.7290896
61	896	74153	55317.111	18835.8889	25.4013848
54	841	72374	51921.530	20452.4705	28.2594170
48	647	54651	39944.387	14706.6128	26.9100526
72	647	43708	39944.387	3763.6128	8.6108100
34	621	64551	38339.203	26211.7969	40.6063374
22	620	69843	38277.465	31565.5347	45.1949869
1	600	37369	37042.708	326.2917	0.8731614
33	572	58984	35314.049	23669.9514	40.1294442
25	513	61967	31671.516	30295.4844	48.8897064
41	460	51580	28399.410	23180.5903	44.9410436
45	450	45040	27782.031	17257.9688	38.3169821
51	448	46605	27658.556	18946.4444	40.6532442
67	428	35537	26423.799	9113.2014	25.6442620

Table 1: State Doctoral and Respondant Counts, and Estimates

STATEICP	Actual Doctoral Count	Total Respondent	Estimated Respondent Count	Difference	% Difference
66	350	20243	21608.247	-1365.2465	-6.7442895
32	321	29940	19817.849	10122.1510	33.8081197
98	311	6718	19200.470	-12482.4705	-185.8063484
65	282	30749	17410.073	13338.9271	43.3800354
53	281	39445	17348.335	22096.6649	56.0189249
46	263	29796	16237.054	13558.9462	45.5059276
31	258	33586	15928.365	17657.6354	52.5743924
42	251	31288	15496.200	15791.8003	50.4723867
4	244	14077	15064.035	-987.0347	-7.0116838
82	214	14995	13211.899	1783.1007	11.8913017
5	177	10401	10927.599	-526.5990	-5.0629647
63	175	19884	10804.123	9079.8767	45.6642363
2	165	14523	10186.745	4336.2552	29.8578476
56	159	18135	9816.318	8318.6823	45.8708701
35	153	19989	9445.891	10543.1094	52.7445564
11	152	9641	9384.153	256.8472	2.6641139
6	131	6860	8087.658	-1227.6580	-17.8958890
64	113	11116	6976.377	4139.6233	37.2402237
68	72	5962	4445.125	1516.8750	25.4423851
37	71	9296	4383.387	4912.6128	52.8465237
36	60	8107	3704.271	4402.7292	54.3077484
81	51	6972	3148.630	3823.3698	54.8389241

Table 2 shows the summary statistics of Table 1, highlighting a mean difference of 19.56%.

Table 2: Laplace Estimation Summary Statistics

Mean Difference	Median Difference	Mean Percent Difference	Median Percent Difference
12785.01	10122.15	19.55614	28.25942

4 Explanation of Differences

Our estimates using the Laplace ratio estimator show some notable differences from the actual numbers of respondents in each state. Here are the key points to consider:

- Magnitude of differences: On average, our estimates differed from the actual numbers by about 12,785 respondents (mean difference), with a median difference of 10,122. This suggests that while some states had larger discrepancies, the typical difference was around 10,000 respondents.
- Variation in education levels: The primary reason for these differences is likely the variation in educational attainment across states. Our method assumed a constant ratio of doctoral degree holders to total population based on California's data. However, this ratio almost certainly varies between states due to differences in economic structures, presence of research institutions, and demographic compositions.

These findings highlight the limitations of applying a single ratio estimator across diverse populations and emphasize the need for more nuanced approaches when estimating population parameters across different regions.