State-Level Respondent Analysis Using Ratio Estimators*

Insights from ACS Data and Methodological Reflections

Jiaxuan Song, Zien Gao and Shuheng (Jack) Zhou

November 21, 2024

Instructions for Obtaining the Data

To download the 2022 ACS PUMS dataset, follow these steps:

1. Visit the IPUMS USA website: [https://usa.ipums.org/usa/%5D(https://usa.ipums.org/usa/).
2. Create an account or log in if you already have one.
3. Navigate to "SELECT SAMPLES", then select ACS data for 2022, click "SUBMIT SAMPLE ELECTIONS"
4. Under "HOUSEHOLD" drop-down tab, select "GEOGRAPHIC", then add "STATEICP" to data cart.
5. Under "PERSON" drop-down tab, select "SEX" under "DEMOGRAPHIC", and "EDUC" under "EDUCATION", add both variables to data cart.
6. View data cart, then click "CREATE DATA EXTRACT", then "SUBMIT EXTRACT"
7. Wait for the data to process then download to local folder.

Make sure to store the dataset in your working directory as `"usa_00004.csv"`.

*The GitHub Repository containing all data, R code, and other files used in this project is located here:https://github.com/Shuhengzhou03/IPUMS_2022.git

Overview of Ratio Estimators Approach

The **ratio estimator** approach is a statistical technique used when you have a sample from which a ratio of two variables can be calculated, and you apply this ratio to the population. In our case, we use California as a reference because we know the total number of respondents in California. By calculating the ratio of doctoral degree holders to total respondents in California, we can estimate the total population of respondents in other states based on the number of doctoral degree holders in those states.

The formula we use is as follows:

$$\text{Estimated Total Respondents in State} = \frac{\text{Doctoral Count in State}}{\text{Doctoral/Total Ratio for California}}$$

Code and Estimates

Below is the R code used to calculate the ratio estimator and compare it to the actual number of respondents in each state.

```
[1] 516430
```

Table 1: Comparison of Estimated vs Actual Respondents by State

State Code	Doctoral Count	Estimated Total Respondents	Actual Total Respondents	Difference	Percentage Error
1	49333	37367	3626205	-3588838	-99
2	15786	11957	1385340	-1373383	-99
3	169872	128670	6981974	-6853304	-98
4	21948	16625	1395231	-1378606	-99
5	15986	12109	1093734	-1081625	-99
6	11300	8559	647064	-638505	-99
11	14619	11073	1018396	-1007323	-99
12	118843	90018	9261699	-9171681	-99

State Code	Doctoral Count	Estimated Total Respondents	Actual Total Respondents	Difference	Percentage Error
13	258461	195772	19677151	-19481379	-99
14	159941	121148	12972008	-12850860	-99
21	142359	107830	12582032	-12474202	-99
22	58574	44367	6833037	-6788670	-99
23	94057	71244	10034118	-9962874	-99
24	105825	80157	11756058	-11675901	-99
25	52211	39547	5892539	-5852992	-99
31	29106	22046	3200517	-3178471	-99
32	33055	25038	2937150	-2912112	-99
33	64989	49226	5717184	-5667958	-99
34	59904	45374	6177957	-6132583	-99
35	17938	13587	1967923	-1954336	-99
36	4892	3705	779261	-775556	-100
37	7887	5974	909824	-903850	-99
40	130985	99215	8683619	-8584404	-99
41	42691	32336	5074296	-5041960	-99
42	24333	18431	3045637	-3027206	-99
43	235306	178233	22244823	-22066590	-99
44	119830	90765	10912876	-10822111	-99
45	37153	28142	4590241	-4562099	-99
46	22308	16897	2940057	-2923160	-99

State Code	Doctoral Count	Estimated Total Respondents	Actual Total Respondents	Difference	Percentage Error
47	120933	91601	10698973	-10607372	-99
48	54520	41296	5282634	-5241338	-99
49	266133	201583	30029572	-29827989	-99
51	36243	27452	4512310	-4484858	-99
52	127642	96683	6164660	-6067977	-98
53	33099	25071	4019800	-3994729	-99
54	71377	54065	7051339	-6997274	-99
56	13533	10251	1775156	-1764905	-99
61	74315	56290	7359197	-7302907	-99
62	83302	63097	5839926	-5776829	-99
63	16181	12256	1939033	-1926777	-99
64	11431	8658	1122867	-1114209	-99
65	23531	17824	3177772	-3159948	-99
66	31316	23720	2113344	-2089624	-99
67	39531	29943	3380800	-3350857	-99
68	7458	5649	581381	-575732	-99
71	516430	391171	39029342	-38638171	-99
72	55444	41996	4240137	-4198141	-99
73	104364	79051	7785786	-7706735	-99
81	6446	4883	733583	-728700	-99
82	18341	13892	1440196	-1426304	-99

State Code	Doctoral Count	Estimated Total Respondents	Actual Total Respondents	Difference	Percentage Error
98	24974	18917	671803	-652886	-97

Explanation of Differences

The estimates derived from the ratio estimator approach will differ from the actual respondent counts due to:

1. Variability in educational attainment across states
2. Differences in the sampling weights used in the ACS dataset
3. Population size and unique demographic characteristics across states

Data Citation

The analysis uses the **2022 ACS Public Use Microdata Sample (PUMS)** dataset, provided by IPUMS USA. The dataset can be accessed at https://usa.ipums.org/usa/. The extracted data includes state-level respondent counts, sex, and educational attainment.

Proper acknowledgment of IPUMS USA is included in this analysis, following their citation requirements (USA 2024)

Code and Tools

This analysis is implemented in **R** using the following key libraries:

-`dplyr`(Wickham et al. 2023) for data manipulation.

-`tibble`(Müller and Wickham 2023) for tidy data representation.

-`readr`(Wickham, Hester, and RStudio 2023) for reading and handling CSV files.

The code provided is fully reproducible and included throughout this `.qmd` file. Users can adapt the methodology or replicate the results by running the Quarto document in RStudio or a similar environment.

Müller, Kirill, and Hadley Wickham. 2023. *Tibble: Simple Data Frames*. https://CRAN.R-project.org/package=tibble.

USA, IPUMS. 2024. "Integrated Public Use Microdata Series, USA: Version 2024." https://usa.ipums.org/usa/.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, Jim Hester, and RStudio. 2023. *Readr: Read Rectangular Text Data*. https://CRAN.R-project.org/package=readr.