Estimating Respondents by State of 2022 ACS USA*

Ziqi Zhu Yuanchen Miao

November 21, 2024

Table of contents

1	Introduction	1
2	Obtain Data	2
3	Overview of ratio estimators approach	2
4	Our estimates and the actual number of respondents.	3
5	Possible reasons of the difference	4
Re	eferences	6

1 Introduction

The R programming language (R Core Team 2023), dplyr (Wickham et al. 2023), and tibble (Müller and Wickham 2023) were used to estimate data and generate tables, package knitr (Xie (2023)) and kabble (Zhu 2024) is used for adding caption to table. The data were gathered from (USA n.d.).

^{*}Code and data are available at: https://github.com/zzq20010617/2022ACS

2 Obtain Data

The data is obtained from the IPUMS USA(USA n.d.), the sample of 2022 only is selected from all of the samples, and "GEOGRAPHIC" and "EDUCATION" are selected in harmonized variables to get variables of "STATEICP" from "GEOGRAPHIC" and "EDUCD" from "EDUCATION". By using the code book, we are able to find out which state is the code of stateicp presents and what degree is the code of educd presents. The format of the data downloading from IPUMS USA has been changed to CSV files. This table Table 1 shows the first several rows of the data we use.

Table 1: 2022 ACS respondent state and educated level

STATEICP	EDUCD	
41	71	
41	64	
41	26	

Then we can filter by doctoral degree by EDUCD and get the number for each STATEICP as Table 2 shows, the number of California (STATEICP=71) shows in Table 3

Table 2: doctoral respondents number of each state in USA 2022

STATEICP	Doctoral Respondents
1	600
2	165
3	2014
4	244
5	177

Table 3: Doctoral Respondents number in California 2022

STATEICP	Doctoral Respondents
71	6336

3 Overview of ratio estimators approach

The ratio estimator is found by first filtering the doctor degree out from all of the respondents by reducing all the other rows which do not have the code of "EDUCD" equal to 116 (where

116 is the doctor degree according to the codebook of IPUMS USA) then we use the method group_by STATEICP to gather the number of doctor degrees together with the same stateicp. Then, by checking the code book, we find the row with stateicp equal to 71, California, and divide the number of doctor degrees in California with the total number of respondents given by Rohan to obtain the ratio estimators

4 Our estimates and the actual number of respondents.

Our estimates and actual number for all states is shown in following table Table 4

Table 4: Estimate and Actual Respondents of Each State in USA (2022)

Actual Total Respondent	Estimated Total Respondents	Doctoral Respondents	State (ICP Code)
3736	37042.708	600	1
14523	10186.745	165	2
7307	124340.024	2014	3
1407	15064.035	244	4
1040	10927.599	177	5
686	8087.658	131	6
964	9384.153	152	11
9316	88779.024	1438	12
20389	174656.370	2829	13
13260	100015.312	1620	14
12804	89952.043	1457	21
6984	38277.465	620	22
10151	61182.207	991	23
12066	74888.009	1213	24
6196	31671.516	513	25
3358	15928.365	258	31
2994	19817.849	321	32
5898	35314.049	572	33
6455	38339.203	621	34
1998	9445.891	153	35
810	3704.271	60	36
929	4383.387	71	37
8876	94520.644	1531	40
51586	28399.410	460	41
3128	15496.200	251	42
217799	168606.061	2731	43
10934	89581.616	1451	44
4504	27782.031	450	45
2979	16237.054	263	46

47	1421	87729.481	109230
48	647	39944.387	54651
49	3216	198548.917	292919
51	448	27658.556	46605
52	1608	99274.458	62442
53	281	17348.335	39445
54	841	51921.530	72374
56	159	9816.318	18135
61	896	55317.111	74153
62	1031	63651.720	59841
63	175	10804.123	19884
64	113	6976.377	11116
65	282	17410.073	30749
66	350	21608.247	20243
67	428	26423.799	35537
68	72	4445.125	5962
71	6336	391171.000	391171
72	647	39944.387	43708
73	1195	73776.727	80818
81	51	3148.630	6972
82	214	13211.899	14995
98	311	19200.470	6718

5 Possible reasons of the difference

The primary reason for discrepancies is the varying levels of educational attainment across states. States with more urban areas and a higher concentration of educational institutions often have a greater proportion of respondents with doctoral degrees compared to California. In such cases, applying California's ratio will result in an overestimation of the total number of doctoral degree holders, like Massachusetts Table 5. Conversely, states with fewer educational institutions or lower educational attainment levels will produce an underestimate.

Another contributing factor is population composition. Differences in population size and demographics can affect the distribution of educational attainment. For example, states with a larger number of immigrants or international students may have more individuals pursuing doctoral degrees. Additionally, states with older populations are more likely to have a higher ratio of residents holding doctoral degrees, as advanced degrees are often obtained later in life.

Table 5: Doctoral respondents, and estimate/actual respondents in Massachusetts 2022

	State (ICP Code)	Doctoral Respondents	Estimated Total Respondents	Actual Total Respondents
3	3	2014	124340	73077

References

- Müller, Kirill, and Hadley Wickham. 2023. *Tibble: Simple Data Frames.* https://CRAN.R-project.org/package=tibble.
- R Core Team. 2023. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.
- USA, IPUMS. n.d. "Variables by Group." University of Minnesota. https://usa.ipums.org/usa-action/variables/group%7D.
- Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. Dplyr: A Grammar of Data Manipulation. https://dplyr.tidyverse.org.
- Xie, Yihui. 2023. Knitr: A General-Purpose Package for Dynamic Report Generation in r. https://CRAN.R-project.org/package=knitr.
- Zhu, Hao. 2024. kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax. https://CRAN.R-project.org/package=kableExtra.