

Predicting Respondants from ACS

2024-10-03

Instructions on how to obtain the data.

1. Go to <https://usa.ipums.org/usa/>
2. Click on “Get Data” under “CREATE YOUR CUSTOM DATA SET”
3. Click on SELECT SAMPLES
4. Remove Defaults and select 2022 ACS
5. Submit sample selections
6. Under Household select Geographic
7. Select STATEICP
8. Under Person select Demograph aic
9. Select Sex
10. Under Person select Education
11. Select EDUC
12. Select View Cart
13. Create Data Extract
14. Confirm all extract requests are as desired.
15. Submit Extract

Brief Overview of the Ratio Estimator Approach

We were provided with the total number of respondents for the state of California. By filtering the data by state and education level, we determined the number of individuals with doctoral degrees in California.

Using this information, we applied the Pierre Laplace’s ratio estimator to estimate the total number of respondents for any other state, X. The formula used is:

$$\text{Total respondents in state X} = \left(\frac{\text{Number of doctoral degrees in state X}}{\text{Number of doctoral degrees in California}} \right) \times \text{Total respondents in California}$$

This method assumes that the ratio of doctoral degree holders to the total population is consistent across states, allowing us to estimate the total respondents for state X based on its doctoral degree count and California's known values.

Your estimates and the actual number of respondents.

	STATEICP	doctoral_degree	predicted_respondents	actual_respondents
1	1	600	37042.708	37369
2	2	165	10186.745	14523
3	3	2014	124340.024	73077
4	4	244	15064.035	14077
5	5	177	10927.599	10401
6	6	131	8087.658	6860
7	11	152	9384.153	9641
8	12	1438	88779.024	93166
9	13	2829	174656.370	203891
10	14	1620	100015.312	132605
11	21	1457	89952.043	128046
12	22	620	38277.465	69843
13	23	991	61182.207	101512
14	24	1213	74888.009	120666
15	25	513	31671.516	61967
16	31	258	15928.365	33586
17	32	321	19817.849	29940
18	33	572	35314.049	58984
19	34	621	38339.203	64551
20	35	153	9445.891	19989
21	36	60	3704.271	8107
22	37	71	4383.387	9296
23	40	1531	94520.644	88761
24	41	460	28399.410	51580
25	42	251	15496.200	31288
26	43	2731	168606.061	217799
27	44	1451	89581.616	109349
28	45	450	27782.031	45040
29	46	263	16237.054	29796
30	47	1421	87729.481	109230
31	48	647	39944.387	54651
32	49	3216	198548.917	292919
33	51	448	27658.556	46605
34	52	1608	99274.458	62442
35	53	281	17348.335	39445

36	54	841	51921.530	72374
37	56	159	9816.318	18135
38	61	896	55317.111	74153
39	62	1031	63651.720	59841
40	63	175	10804.123	19884
41	64	113	6976.377	11116
42	65	282	17410.073	30749
43	66	350	21608.247	20243
44	67	428	26423.799	35537
45	68	72	4445.125	5962
46	71	6336	391171.000	391171
47	72	647	39944.387	43708
48	73	1195	73776.727	80818
49	81	51	3148.630	6972
50	82	214	13211.899	14995
51	98	311	19200.470	6718

Some explanation of why you think they are different.

The difference between the estimated and actual number of respondents arises from several key factors. The ratio estimator assumes that the proportion of doctoral degree holders in California reflects other states, but educational attainment varies. States with more research institutions, like Massachusetts, often have more residents with advanced degrees compared to others, which the estimator overlooks.

Additionally, population size differences matter. Larger states, such as Texas, have more respondents, which the ratio estimator doesn't fully capture. Sampling biases in survey response rates across states can also lead to discrepancies, as some states may be over- or underrepresented in the sample.

Finally, socioeconomic disparities affect educational achievement, with wealthier states often having more doctoral degree holders. The estimator's assumption of uniformity across states doesn't consider these variations, leading to the observed differences between estimated and actual respondent numbers.