Community Survey 2022 Ratio Estimator Analysis*

Community-based paramedic service strategies that focus on medical emergencies, emergency transfers, and vehicle accidents may help to effectively manage EMS demand

Aamishi, Daniel, Dennis, Vandan - Group 35

October 3, 2024

date-format: long

Instructions on How to Obtain the Data The following is the instructions on how to obtain the data,

Go to IPUMS https://usa.ipums.org/usa/

Then click "get data"

Click select samples

Unselect Default samples from each year and select ACS 2022. Then click samle selctions

Then select vaiable by using the search function under harmonized variables to search for STATEICP and EDUC

Once both are added go to cart

Click create data extract. Change the data format to csv. Submit extract. (You may at this point have to log in to your account or create a new account)

Download the csv.

An Overview of The Ratio Estimator A ratio estimator is a statistical technique used in survey sampling to estimate a population parameter (such as the total or mean) by leveraging auxiliary information that is correlated with the variable of interest. This method improves the precision of the estimate by using the known total of an auxiliary variable that is easier to measure or has a known population total.

Key Elements

*Code and data are available at: https://github.com/vandanppatel/week5exercise

Key Concepts: Using the Ratio of Two Variables: A ratio estimator relies on the assumption that two variables, say "Y" (the variable you want to estimate) and "X" (an auxiliary variable), are linearly related. "X" should be correlated with "Y", and the total or mean of "X" should be known.

The ratio estimator for the population mean of "Y" is calculated as the sample mean of "Y" divided by the sample mean of "X", multiplied by the known population mean of "X".

Purpose: The main idea is to use an auxiliary variable that is easy to measure and has a known mean or total. If "X" is related to "Y", this can give you a more precise estimate of the population mean or total of "Y".

Efficiency: Ratio estimators often perform better than simply using the sample mean, especially when "Y" and "X" are highly correlated. They can reduce the variability of your estimate, making it more precise.

Bias: Ratio estimators may introduce some bias (meaning the estimate could be systematically off), but this bias tends to decrease as the sample size grows. For large samples, the ratio estimator usually gives reliable results.

As shown in Table 1, column estimated_value represents estimated number of doctoral respondents, column actual_number represents actual number of doctoral respondents.

Table 1: The Number of Actual and Estimated People in Each State with a Doctoral Degree

STATEICP	total_count	estimated_value	actual_number
1	37369	605.28512	600
2	14523	235.23658	165
3	73077	1183.66615	2014
4	14077	228.01249	244
5	10401	168.47040	177
6	6860	111.11499	131
11	9641	156.16029	152
12	93166	1509.05813	1438
13	203891	3302.52850	2829
14	132605	2147.87211	1620
21	128046	2074.02761	1457
22	69843	1131.28337	620
23	101512	1644.24263	991
24	120666	1954.48992	1213
25	61967	1003.71171	513
31	33586	544.00990	258
32	29940	484.95374	321
33	58984	955.39451	572
34	64551	1045.56610	621

Table 1: The Number of Actual and Estimated People in Each State with a Doctoral Degree

STATEICP	total_count	estimated_value	actual_number
35	19989	323.77222	153
36	8107	131.31329	60
37	9296	150.57214	71
40	88761	1437.70805	1531
41	51580	835.46807	460
42	31288	506.78800	251
43	217799	3527.80361	2731
44	109349	1771.18259	1451
45	45040	729.53629	450
46	29796	482.62130	263
47	109230	1769.25508	1421
48	54651	885.21065	647
49	292919	4744.56129	3216
51	46605	754.88541	448
52	62442	1011.40553	1608
53	39445	638.91117	281
54	72374	1172.27929	841
56	18135	293.74202	159
61	74153	1201.09468	896
62	59841	969.27578	1031
63	19884	322.07148	175
64	11116	180.05163	113
65	30749	498.05753	282
66	20243	327.88639	350
67	35537	575.61126	428
68	5962	96.56961	72
71	391171	6336.00000	6336
72	43708	707.96119	647
73	80818	1309.05115	1195
81	6972	112.92911	51
82	14995	242.88181	214
98	6718	108.81494	311

Ratio estimator approach is biased to the order O(1/n), but with hundreds of thousands of observations, it is not very biased but bias still exists. Also since we are using the ratio of doctoral degrees for only California, we are falsely assuming that the ratio is the same for all other states, and doctoral degrees have the same relationship with recipients in California as in other states. Also if doctoral degrees and recipients do not have a high correlation the

estimates will be less accurate. Random chance will also create a difference between estimated vs real numbers.

Citation Steven Ruggles, Sarah Flood, Matthew Sobek, Daniel Backman, Annie Chen, Grace Cooper, Stephanie Richards, Renae Rodgers, and Megan Schouweiler. IPUMS USA: Version 15.0 2022 ACS. Minneapolis, MN: IPUMS, 2024. https://doi.org/10.18128/D010.V15.0 R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

01-usage.txt: https://drive.google.com/file/d/1LTEbl5D05fMIS-MeLb_oihlN16F-WL6t/view?usp=sharing