Analysis of ACS 2022

Lexun Yu   Colin Sihan Yang   Siddharth Gowda
Tanmay Shinde

October 3, 2024

This document provide instructions on downloading 2022 ACS data, a brief overview of the ratio estimators approach, estimates and the actual number of respondents and our explanation of why we think they are different.

1 Introduction

This paper is written with the help of R (R Core Team 2023). All the data are gathered from IPUMS USA (Ruggles et al. 2024). The code sections are done with the help of tidyverse (Wickham et al. 2019), janitor (Firke 2023) and knitr (Xie 2014).

2 Downloading Data

We followed the steps below to download the data:

1. Visit IPUMS USA at: https://usa.ipums.org/

2. Click on "Get Data" under "Create Your Custom Data Set"

3. Use the filter in "Select Harmonized Variables" and select "Geographic" under "Household". Check the "STATEICP" checkbox.

4. Use the filter in "Select Harmonized Variables" and select "Demographic" under "Person". Check the "SEX" checkbox.

5. Use the filter in "Select Harmonized Variables" and select "Education" under "Person". Check the "EDUC" checkbox.

6. Select "Select Samples", check only "ACS" in 2022, and click "Submit Sample Selections"

7. Click "View Cart" on the top of the page, click on "Create Data Extract". Change the data format to "Comma delimited (.csv)" and apply.

8. Submit extract, login and wait for the confirmation email.

3 Estimating the total numner of respondents

The ratio estimator method is a technique used to estimate the population total of interest in one state by leveraging the relationship between two related variables in a reference group.

In our case, we are given information about the number of respondents with doctoral degrees in each state, as shown in **?@tbl-ttl**. We also know the total number of respondents in California (391,171). The ratio between the number of people with doctoral degrees and the total number of respondents in California can be used to estimate the total number of respondents in other state.

We follow these steps to estimate the total number of respondents in all states.

1. The ratio in California is calculated as the number of people with doctoral degrees in California divided by the total number of respondents in California: Ratio in California $= \frac{\text{Doctoral Count in California}}{\text{Total Respondents in California}} = \frac{6336}{391171} \approx 0.0161975$. The result is displayed in "Doctor Proportion" on Line 71 in **?@tbl-ttl**.

2. We can then use this ratio to estimate the total number of respondents in other states. For any state x, the estimated total number of respondents is calculated by: Estimated Total Respondents in State $x = \frac{\text{Doctoral Count in State } x}{\text{Ratio in California}}$. The result is shown in **?@tbl-est**.

The comparison between the estimated and actual data is displayed in **?@tbl-est**.

The table of estimated total respondents by state is shown in Table **??**.

Bibliography

Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data.* https://CRAN.R-project.org/package=janitor.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Ruggles, Steven, Sarah Flood, Matthew Sobek, Daniel Backman, Annie Chen, Grace Cooper, Stephanie Richards, Renae Rodgers, and Megan Schouweiler. 2024. "IPUMS USA: Version 15.0." Minneapolis, MN: IPUMS. https://doi.org/10.18128/D010.V15.0.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC.