Total and estimated number of US respondents that had a doctorate degree as highest level of attainment*

Chris Yong Hong Sen Xuanang Ren Caichen Sun Irene Liu
Mingxuan Liu

October 3, 2024

1 Introduction

We will be analysing how many respondents were there in each state (STATEICP) that had a doctoral degree as their highest educational attainment (EDUC). Particularly, we aim to use Laplace ratio estimators to support our findings.

The remainder of this paper is structured as follows. Section 2 will contain information about data retrieval and data manipulation using R Core Team (2023).

2 Data

2.1 Data Retrieval

Our dataset was obtained from USA Ipums, which is an online US census data for social, economic and health research resource (Ruggles et al. 2024a). Following the instructions outlined by USA Ipums, we are able to select 2022 American Census Dataset, where we subset the state ($STATEICP$) and educational attainment ($EDUCD$) of the individual respondent. These variables are relevant to our research question introduced in Section 1.

*Code and data are available at: https://github.com/Monoji77/USA_data.

Table 1: states with doctoral as educational attainment

```
# A tibble: 51 x 3
   `state ID` `actual doctorate count` `estimated doctorate count`
        <dbl>                    <int>                        <dbl>
 1          1                      600                          605
 2          2                      165                          235
 3          3                     2014                         1184
 4          4                      244                          228
 5          5                      177                          168
 6          6                      131                          111
 7         11                      152                          156
 8         12                     1438                         1509
 9         13                     2829                         3303
10         14                     1620                         2148
# i 41 more rows
```

2.2 Ratio Estimators Approach Overview

As for data manipulation, we use tidyverse library from R and we can obtain the total number of respondents in the state of california (R Core Team 2023). Furthermore, we can also obtain the total number of respondents in California with a doctorate degree by finding EDUCD value of 116. We were able to find this value in the documentation provided by USA Ipums (Ruggles et al. 2024b). Finally the laplace ratio can be calculated as…

californian_doctorates_total_respondents / californian_total_respondents

We then use this value obtained and multiply it by the number of respondents in other states to determine an estimate number of respondents in that state with a doctorate degree. Table 1 show us the top few entries of actual number of doctorates per US state and estimated number of doctorates per US state using the aforementioned laplace ratio. Note that the estimated values could differ since the ratio between doctorate respondents and total respondents in california could be different from that of another state. One reason this could be is that there are more universities in some states compared to california (vice versa for states with lesser universities). There are other socio-demographic related reasons as well such as gender and family's household income. This is reflected in Table 1 where we can observe how the difference between estimated and actual total number of doctorate respondents from the first observation only differs by 0.8% while that of the second observation had a larger difference of 42%.

References

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. 2024a. *U.s. CENSUS DATA FOR SOCIAL, ECONOMIC, AND HEALTH RESEARCH*. https://usa.ipums.org/usa/.

———. 2024b. *U.s. CENSUS DATA FOR SOCIAL, ECONOMIC, AND HEALTH RESEARCH*.https://usa.ipums.org/usa-action/variables/EDUC#description_section.