Estimation of Respondents Using Ratio Estimators Approach

Yihang Xu

2024-10-03

Introduction

This report uses the ratio estimators approach to estimate the total number of respondents in each state based on the number of respondents with doctoral degrees. We compare these estimates with the actual number of respondents and provide insights into why the estimates may differ from the actual values.

Instructions on How to Obtain the Data

The dataset used for this exercise is located in `/Users/xuyihang/Downloads/usa_00001.csv`. It contains information about individuals across various states, including their educational attainment. To load the dataset into R, we used the following code:

Ensure that the dataset has columns `STATEICP` for state identification and `EDUC` for educational attainment. The values for doctoral degrees should be filtered from the `EDUC` column.

Counting Respondents with Doctoral Degrees by State

Using the codebook and the dataset, we can count how many respondents in each state had a doctoral degree as their highest educational attainment. Below is the code that creates a tibble with this information:

```
# A tibble: 51 x 2
   STATEICP doctoral_count
      <int>          <int>
 1        1            600
```

```
 2        2            165
 3        3           2014
 4        4            244
 5        5            177
 6        6            131
 7       11            152
 8       12           1438
 9       13           2829
10       14           1620
# i 41 more rows
```

This tibble shows the number of respondents with doctoral degrees in each state.

Estimation and Actual Numbers

Below is the code used to estimate the total number of respondents in each state and compare it with the actual number of respondents:

```
# A tibble: 51 x 3
   STATEICP actual_total estimated_total
      <int>        <int>           <dbl>
 1        1        37369          37043.
 2        2        14523          10187.
 3        3        73077         124340.
 4        4        14077          15064.
 5        5        10401          10928.
 6        6         6860           8088.
 7       11         9641           9384.
 8       12        93166          88779.
 9       13       203891         174656.
10       14       132605         100015.
# i 41 more rows
```

Discussion

The ratio estimators approach assumes that the ratio of doctoral degrees to the total population remains relatively stable across states. However, differences between the estimated and actual numbers can arise due to various reasons:

1. **Variability in Educational Attainment**: Different states may have different proportions of individuals with doctoral degrees compared to California. For example, states with major research universities might have disproportionately higher numbers of doctoral degree holders, while others may have lower proportions.
2. **Population Structure**: The age, socioeconomic, and occupational structure of each state's population can influence educational attainment. States with younger populations or economies focused on industries that do not require advanced degrees may have lower numbers of individuals with doctoral degrees.
3. **Sampling and Data Limitations**: The dataset might have some sampling or reporting limitations, which can affect the accuracy of our estimates.

These differences highlight the limitations of applying a fixed ratio from one state to others, especially when the demographic and educational characteristics differ.

Conclusion

The ratio estimators approach provided a useful method for estimating the total number of respondents in each state. However, the comparison with actual numbers revealed notable differences, suggesting that further refinement is needed to account for state-specific characteristics. Future work could involve using more localized ratios or additional variables to improve the accuracy of estimates.