

Reflection 4

Jin Zhang

The number of the respondents in each state (STATEICP) that had a doctoral degree as their highest educational attainment (EDUC):

```
# A tibble: 51 x 2
  STATEICP doctoral_count
  <dbl>      <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
```

Obtain the data.

To access the data from IPUMS USA, we initiated the process by navigating to the IPUMS website and selecting “IPUMS USA.” We proceeded by clicking “Get Data” and chose the “2022 ACS” under “SELECT SAMPLE.” For state-level data, we selected “HOUSEHOLD” then “GEOGRAPHIC” and added “STATEICP” to our cart. For data on individuals, we clicked on the “PERSON” category and added “EDUC” to our selection. We reviewed our choices by clicking on “VIEW CART” followed by “CREATE DATA EXTRACT.” We altered the “DATA FORMAT” to “.dta” and submitted the extract. After either logging in or signing up, we were notified via email once the extract was prepared. We downloaded the file, named “usa_00004.dta,” and saved it on our local machine for analysis in R.

Ratio estimators approach.

The ratio estimation method is a statistical technique that uses ratios derived from a sample to calculate a population total or average. This method involves determining the ratio of a particular attribute (e.g., the number of people with doctoral degrees) to the overall population of a known group (e.g., the State of California). This ratio is then used to estimate the total number of people in other groups, assuming a similar pattern exists in the population in general.

Estimates and the actual number of respondents.

```
# A tibble: 51 x 3
  STATEICP actual_total estimated_total
  <dbl>      <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
```

Some explanation of why you think they are different.

The estimated total number of respondents in each state using the ratio estimators approach might differ from the actual numbers for these reasons:

1. **Diverse Populations:** Different states have their own unique demographic and educational characteristics. A ratio calculated from one state, such as California, may not be accurate in the other states with different characteristics.
2. **Sampling Errors:** The ratio is derived from a sample of the population rather than the entire true population. This can introduce errors because the sample might not perfectly represent the larger group.

3. Non-Representative Samples: If the sample used to calculate the ratio comes from a state that is not typical of other states, the ratio might not be suitable for estimating numbers in other states.
4. Uniformity Assumption: This method assumes that the conditions in one state are similar to those in others, which often is not the case. Differences in education systems, economies, and cultures can make this assumption incorrect.