2022ACS

Siyuan Lu

2024-10-03

Obtain the data

We get the data from IPUMS USA. We first select samples 2022 ACS. Next we select geographic household variable STATEICP, demographic person variable SEX, and education person variable EDUC. Next, we create the data extraction and login to IPUMS account. Finally, download the data usa_00001.csv.gz and gunzip in R.

Ratio estimators approach

The ratio estimator is a statistical method used to improve the precision of estimates by leveraging the correlation between a target variable (Y) and an auxiliary variable (X) associated with Y. It assumes a linear relationship between Y and X, or that their ratio remains constant. Ratio estimators are particularly effective when the auxiliary variable is easily measured and highly correlated with the target variable.

Estimates and the actual number of respondents

```
# A tibble: 51 x 3
   STATEICP actual_respondents estimated_respondents$STATEICP $doctoral
      <dbl>              <int>                           <dbl>     <dbl>
 1        1              37369                         3305670.    10187.
 2        2              14523                         3305670.    10187.
 3        3              73077                         3305670.    10187.
 4        4              14077                         3305670.    10187.
 5        5              10401                         3305670.    10187.
 6        6               6860                         3305670.    10187.
 7       11               9641                         3305670.    10187.
```

```
 8       12              93166                       3305670.    10187.
 9       13             203891                       3305670.    10187.
10       14             132605                       3305670.    10187.
# i 41 more rows
```

Why they are different

The estimate the total number of respondents in each state is different with the actual number of respondents in each state for following reasons:

The ratio estimator assumes that the proportion of people with doctoral degrees relative to the total population is similar across states. However, education levels vary significantly due to state-specific factors.

The assumption of uniformity in educational attainment is often not valid across states, leading to inaccurate estimates.