week5 Refelction 4

```r
library(readr)
library(tidyverse)
```

```
-- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
v dplyr     1.1.4     v purrr     1.0.2
v forcats   1.0.0     v stringr   1.5.1
v ggplot2   3.5.1     v tibble    3.2.1
v lubridate 1.9.3     v tidyr     1.3.1
-- Conflicts ------------------------------------------- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becom
```

```r
library(knitr)
clean <-read_csv("cleaned_data.csv")
```

```
Rows: 51 Columns: 2
-- Column specification -----------------------------------------------------
Delimiter: ","
dbl (2): STATEICP, respondents

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
raw<-read_csv("usa_00001.csv")
```

```
Rows: 3373378 Columns: 14
-- Column specification -----------------------------------------------------
Delimiter: ","
dbl (14): YEAR, SAMPLE, SERIAL, CBSERIAL, HHWT, CLUSTER, STATEICP, STRATA, G...
```

```
i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Instruction

Instructions for Extracting Data from IPUMS USA 1. Access the IPUMS USA Website Visit IPUMS USA. 2. Search for Variables In the search bar, enter the following variables one at a time: EDUC (Education) SEX (Gender) STATEICP (State, using the ICP coding system) 3. Add Variables to Your Cart After searching for each variable, select it and add it to your cart. 4. Create Data Extract Once all desired variables are in your cart, proceed to create the data extract. 5. Change Data Format In the extract request page, select CSV as the data format. This will make it easier for you to work with the data. 6. Submit Your Request Review your selections and submit the data extract request. 7. Download Your Data After your request is processed, navigate to the download page at Download Data. Locate your data file, which will be named usa_00001.csv, and download it.

Overview of Ratio Estimator

First, we start with the how many respondents have a doctor degree as their highest educational attainment for each state. See "Table 1" for overview.

ID	Counts
71	6336
49	3216
13	2829
43	2731
3	2014
14	1620

Table 1: Sample of doctor degree for each state

Then, we find the ratio for California, which has the ID 71 given by codebook. After that we calculate the ratio of doctoral degree respondents to total respondents in California by number of respondents with doctoral degrees in California divided by total number of respondents in California, which is given by the teacher(391,171). Then, use the ratio, we can calculate the number of respondents for every state by number of doctor degree divided by the ratio. See table 2 for overview of the estimated total respondents.

ID	Counts	Estimated
71	6336	391171.0
49	3216	198548.9
13	2829	174656.4
43	2731	168606.1
3	2014	124340.0
14	1620	100015.3

Table 2: Sample of doctor degree for each state

Then we compare with the actual number.

ID	Counts	Estimated	Actual
71	6336	391171.0	391171
49	3216	198548.9	292919
13	2829	174656.4	203891
43	2731	168606.1	217799
3	2014	124340.0	73077
14	1620	100015.3	132605

Table 3: Sample of estimate v.s. actula

Explaination of Difference

The Laplace ratio estimator is based on the ratio of doctoral degree holders in California to the total respondents in California, applied to other states. This method assumes that the proportion of doctoral degree holders is similar across states, which may not always be true. Each state has different demographics, education systems, and labor markets, leading to variations in educational attainment levels. For instance, states with a higher concentration of research institutions or universities may have more doctoral degree holders. If certain respondents in a state did not report their education level, this could lead to undercounts in the actual totals. The estimation assumes that the ratio of doctoral degree holders to total respondents is consistent across states. If certain states have unique educational trends or policies affecting the attainment of doctoral degrees, the estimate may not hold true.