

# Analysis of Doctoral Degree Holders in the 2022 ACS

Feiyang Xue

## Introduction

This report aims to estimate the total number of respondents in each state based on the ratio of respondents with doctoral degrees using the 2022 ACS sample data obtained from IPUMS. We will use California's sample data as the baseline for applying the Laplace ratio estimators approach.

## Data Access

The data was obtained from the IPUMS USA platform. The relevant variables include: - **StateICP**: State identifier - **EDUCD**: Educational attainment, specifically for doctoral degrees

We downloaded the dataset and loaded it into R for analysis. Below is the code used to clean and process the data.

Ratio Estimation Calculation We use the provided sample size of 391,171 respondents from California to calculate the ratio of doctoral degree holders to the total respondents in California. This ratio is then applied to estimate the total number of respondents in other states based on their number of doctoral degree holders.

```
# Subset for California Doctoral Data
California_doctor_data <- subset(Doctor_data_cleaned, State_name == "71 California")

# Calculate the California Ratio
california_ratio <- as.integer(California_doctor_data$Doctoral_Degree_Count) / 391171

# Apply Ratio Estimation to Other States
Total_data_cleaned$Estimated_Total_Respondents <-
```

```

ceiling(as.integer(Doctor_data_cleaned$Doctoral_Degree_Count) / california_ratio)

# Calculate the Difference between Estimated and Actual Respondents
Total_data_cleaned$Difference <- Total_data_cleaned$Total_Count -
  Total_data_cleaned$Estimated_Total_Respondents
head(Total_data_cleaned)

```

	State_name	Total_Count	Estimated_Total_Respondents	Difference
1	1 Connecticut	37369	37043	326
2	2 Maine	14523	10187	4336
3	3 Massachusetts	73077	124341	-51264
4	4 New Hampshire	14077	15065	-988
5	5 Rhode Island	10401	10928	-527
6	6 Vermont	6860	8088	-1228

The table below presents a comparison between the actual total respondents and the estimated total respondents in each state, along with the differences.

```

library(knitr)
kable(head(Total_data_cleaned))

```

State_name	Total_Count	Estimated_Total_Respondents	Difference
1 Connecticut	37369	37043	326
2 Maine	14523	10187	4336
3 Massachusetts	73077	124341	-51264
4 New Hampshire	14077	15065	-988
5 Rhode Island	10401	10928	-527
6 Vermont	6860	8088	-1228

## Analysis

There may be multiple reasons cause the difference; State-Specific Education Distribution: The ratio estimator is sensitive to the distribution of educational attainment within each state. States that have a different ratio of doctoral degree holders to the general population compared to California may exhibit larger differences.

Population Size and Demographics: States with larger populations or a significantly different demographic composition might see bigger discrepancies because the ratio estimator is based on the specific characteristics of California's population, which may not generalize well to other states.

Sampling and Weighting: ACS data uses weights to estimate population figures. If the sample sizes in some states are smaller or if their weighting factors differ significantly from California's, this could lead to over- or under-estimation of the total number of respondents.