

# Homework 2

Research Methods, Spring 2025

Justin Hahm

My answers to the homework questions are described below. Note that I do the analysis for these answers in a separate R script. My analysis file is available in the analysis folder. The GitHub repository for this work is available [here](#). Enjoy!

## Summarize the Data

1. How many hospitals filed more than one report in the same year? Show your answer as a line graph of the number of hospitals over time.

The graph was compiled from a pre-filtered version of the data set. I included all years available from 1997. The duplicate reports over years are illustrated in Figure 1.

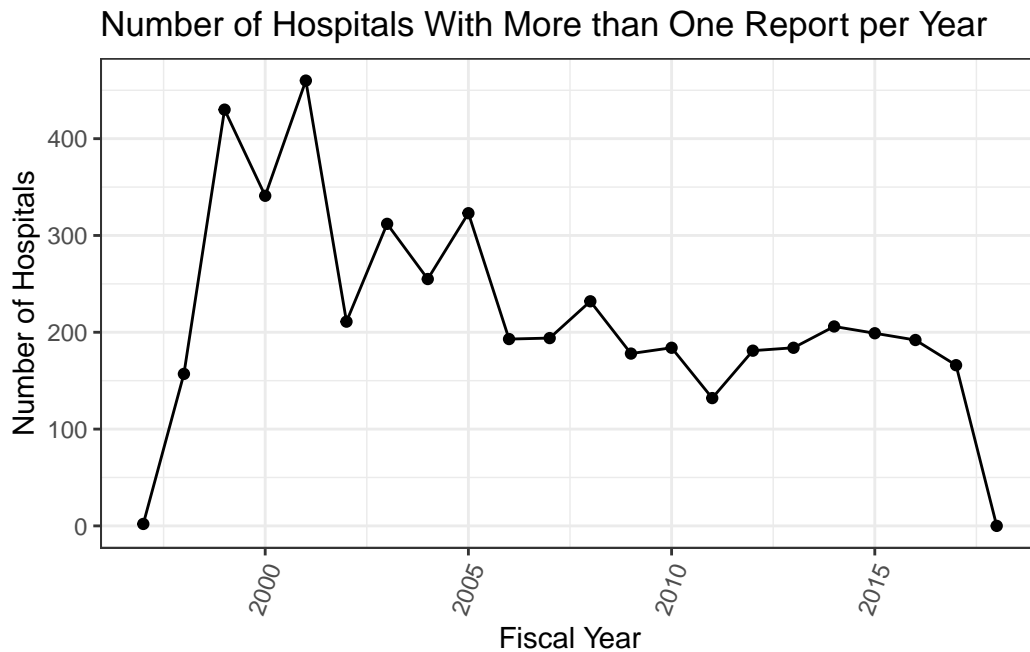


Figure 1: Duplicate over years starting from 1997

2. After removing/combining multiple reports, how many unique hospital IDs (Medicare provider numbers) exist in the data?

After filtering, I was able to include data starting from 1997. See [Figure 2](#).

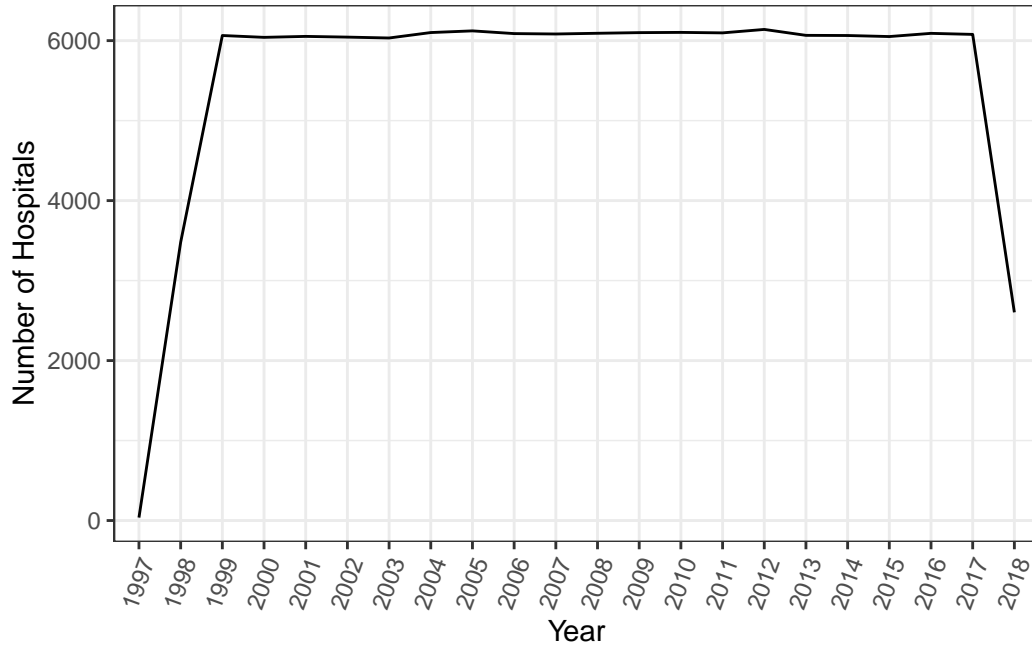


Figure 2: Unique hospital IDs starting from 1997

3. What is the distribution of total charges (tot\_charges in the data) in each year? Show your results with a “violin” plot, with charges on the y-axis and years on the x-axis.

I filtered for the interquartile range to remove outliers in the data. I then performed a log transformation on the total charges. See Figure 3.

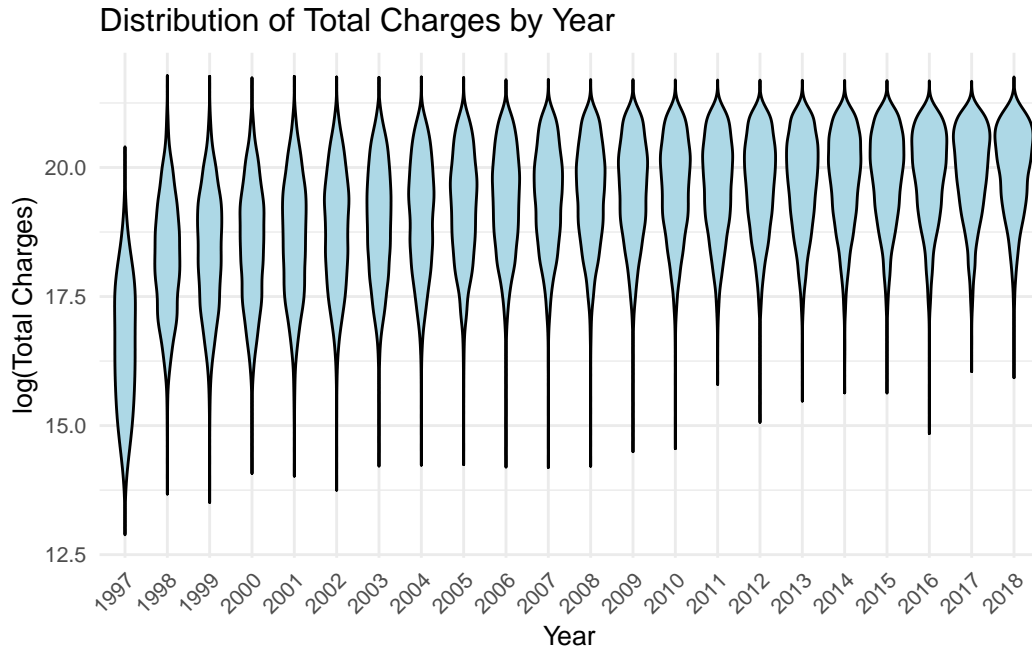


Figure 3: Total charges over years starting from 1997

4. What is the distribution of estimated prices in each year? Again present your results with a violin plot, and recall our formula for estimating prices from class. Be sure to do something about outliers and/or negative prices in the data.

See Figure 4.

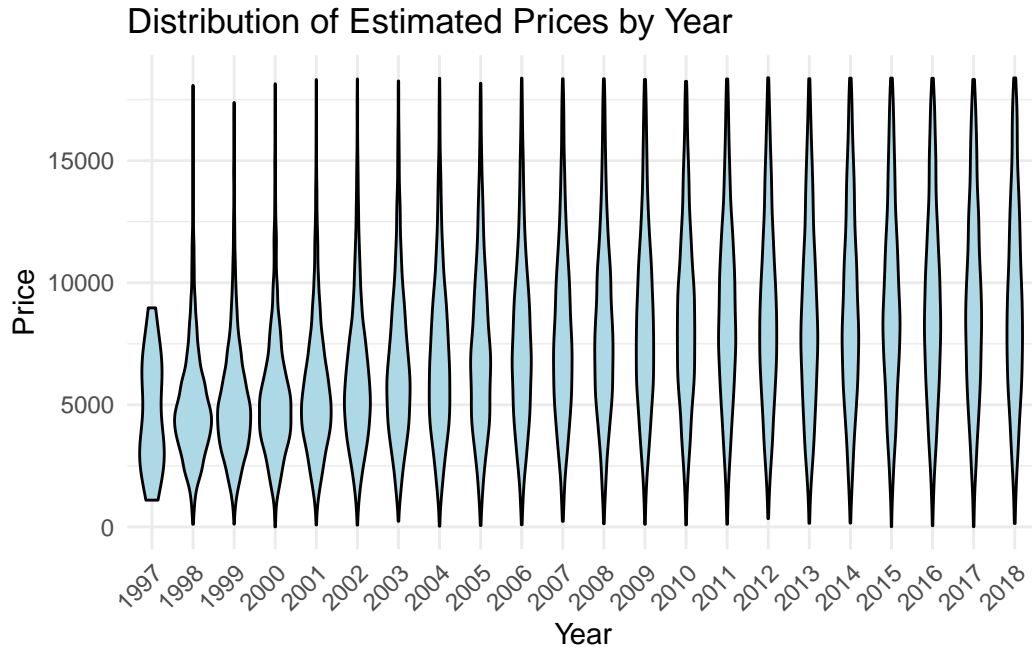


Figure 4: Prices over years starting from 1997

## Estimate ATEs

5. Calculate the average price among penalized versus non-penalized hospitals.

The average price among non-penalized hospitals in 2012 is 9,896.31 and the average price among penalized hospitals is 9,560.41.

6. Split hospitals into quartiles based on bed size. To do this, create 4 new indicator variables, where each variable is set to 1 if the hospital's bed size falls into the relevant quartile. Provide a table of the average price among treated/control groups for each quartile. See Table 1.

Table 1: Quartiles by bed size in 2012

Quartile	Nonpenalty	Penalty
1	7,684	8,319
2	8,511	8,691
3	9,857	10,127
4	12,356	12,068

7. Find the average treatment effect using each of the following estimators, and present your results in a single table: See Table 2.

Table 2: ATE

Method	ATE
Inverse Variance Matching	199.53
Mahalanobis Matching	199.53
IPW	199.53
Regression Adjustment (Penalty=1 vs Penalty=0)	199.53

8. With these different treatment effect estimators, are the results similar, identical, very different?

The results from the different treatment effect estimators are identical.

9. Do you think you've estimated a causal effect of the penalty? Why or why not? (just a couple of sentences)

I believe I have estimated a causal effect of the penalty because I see convergence across different methods. This indicates consistency. The selection on observables assumption should be satisfied by matching within the quartiles.

10. Briefly describe your experience working with these data (just a few sentences). Tell me one thing you learned and one thing that really aggravated or surprised you.

It was difficult to keep track of the different variable names and the dfs after filtering. While working with this dataset, I realized the importance of being organized and working through my code sequentially. The values would change if I ran the code out of order while debugging, which quickly became frustrating.