Homework 2

Research in Health Economics, Spring 2024
https://github.com/riadharnidharka/Econ470hw2

Ria Dharnidharka

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.
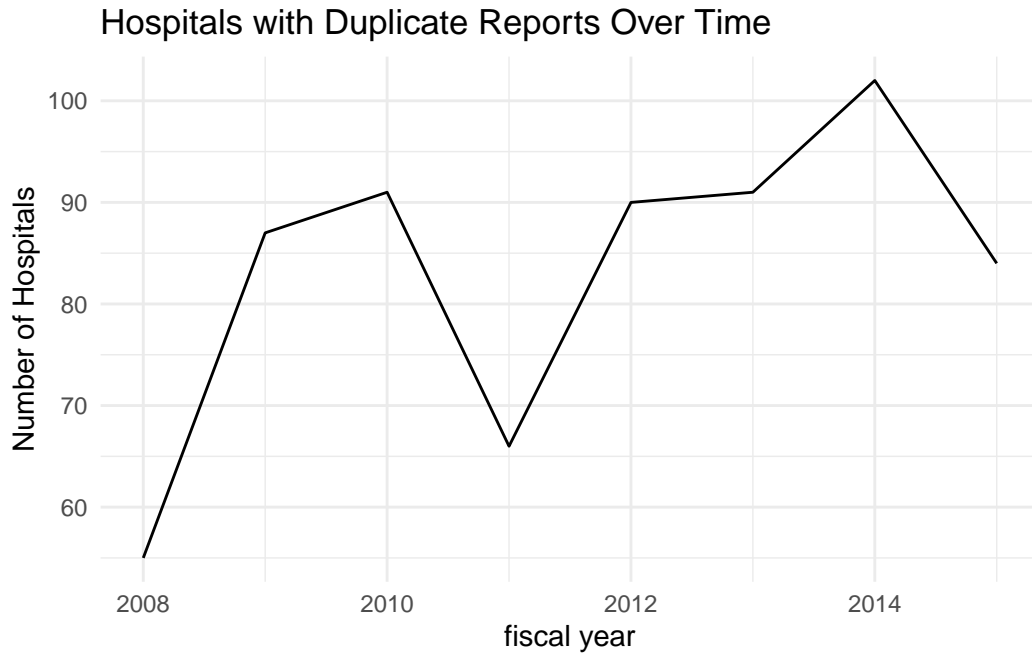

Figure 1: Hospitals with Duplicates

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

There are 6747 unique hospital ids in the dataset (for years 2008-2015).

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. For a nice tutorial on violin plots, look at Violin Plots with ggplot2.
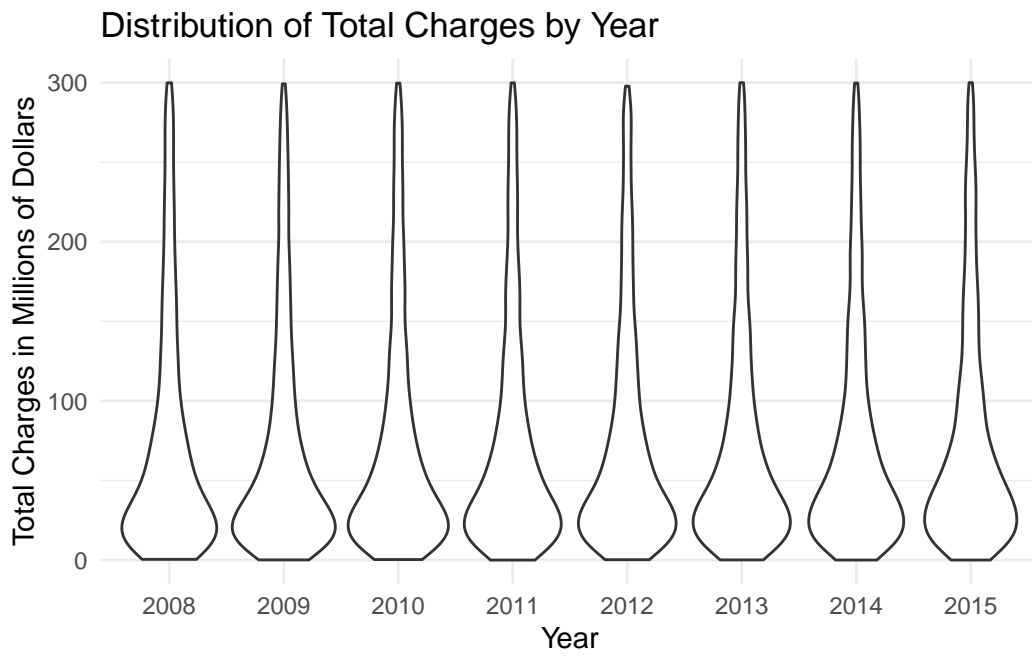

Figure 2: Total Charges

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.
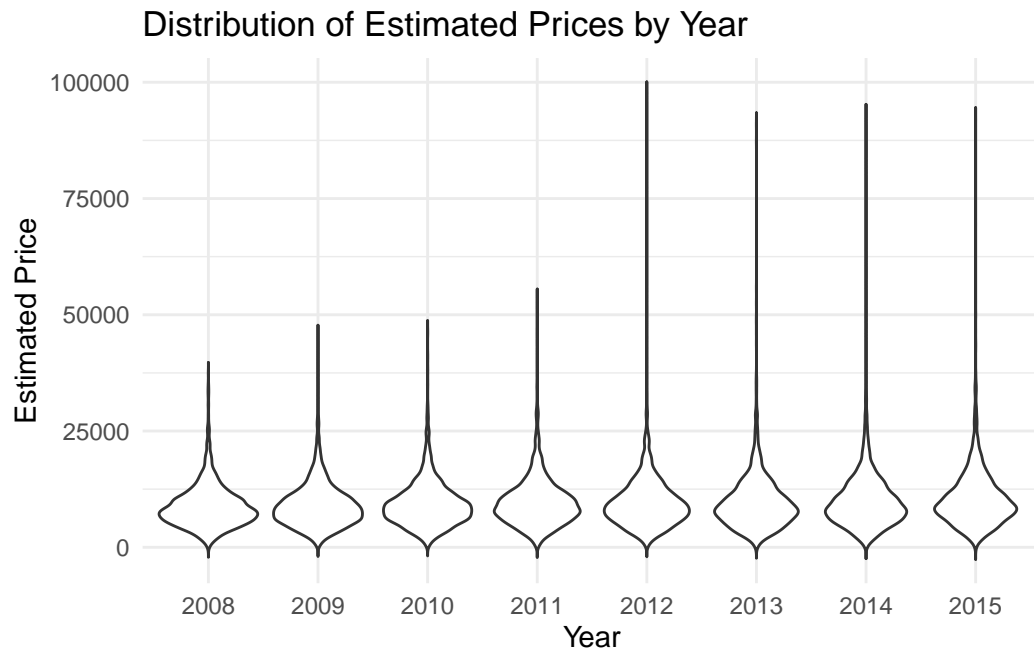

Figure 3: Estimated Prices

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

The average price among non-penalized hospitals is $9560.41, while the average price among penalized hospitals is $9896.31.

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.

Table 1: Quartile Price Summary

Quartile	Treated Avg Price	Non-Treated Avg Price
1	8,318.709	7,684.240
2	8,690.891	8,510.959
3	10,127.130	9,856.928
4	12,068.479	12,355.606

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

- Nearest neighbor matching (1-to-1) with inverse variance distance based on quartiles of bed size
- Nearest neighbor matching (1-to-1) with Mahalanobis distance based on quartiles of bed size
- Inverse propensity weighting, where the propensity scores are based on quartiles of bed size
- Simple linear regression, adjusting for quartiles of bed size using dummy variables and appropriate interactions as discussed in class

Table 2: Average Treatment Effects

Estimator	ATE
Nearest Neighbor (Inverse Variance Distance)	190.1416
Nearest Neighbor (Mahalanobis Distance)	190.1416
Inverse Propensity Weighting	190.1416
Simple Linear Regression	190.1416

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

With these different treatment effect estimators, the results ended up being identical. This may be because they rely on similar assumptions within the data, all account for confounding, and they have similar treatment and control groups.

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

I would guess that we are estimating a causal effect of the penalty because these estimators adjust for confounding variables and create balanced groups, which allows for an estimation of causality. Their purpose is to estimate causality, through minimizing bias and confounding.

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.

My experience working with the data this time around was much better, and I felt a lot more comfortable every step of the way. One thing I learned is how to do the analysis of the data without looking at the exact code, and instead using the slides and chatgbt to guide me in creating my own code. One thing that aggravated me is not knowing if my answers were correct or not between submissions as everyone was in a different place, so it was hard to figure out which answers to revise.