Homework 2

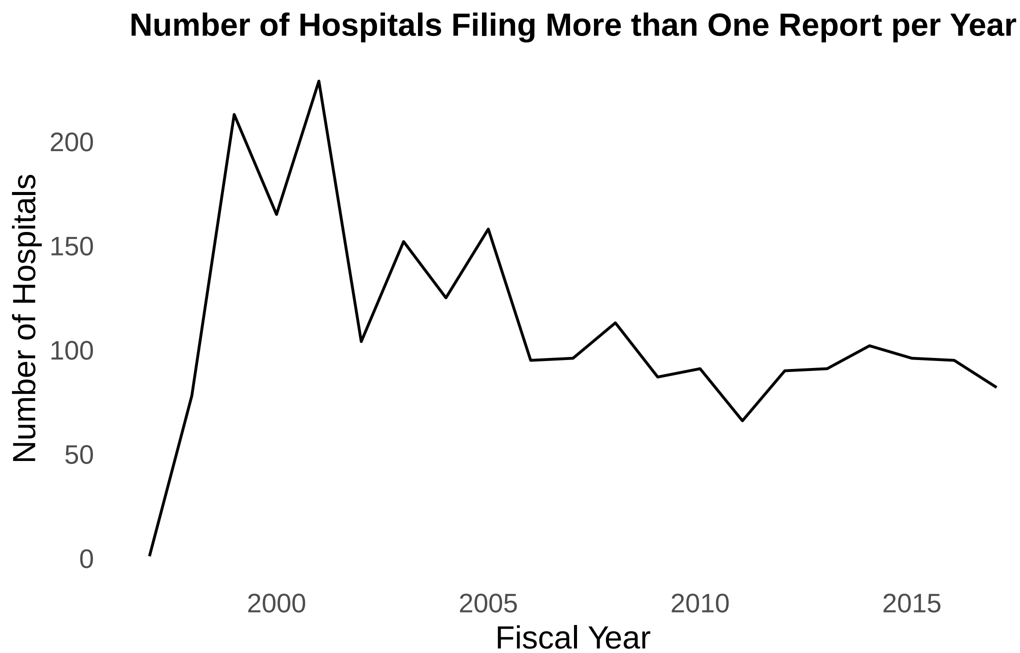
Research in Health Economics, Spring 2025

Lisbeth Vargas

The following is my submission for Homework 2. Note that the setup and analysis for these responses are in a seperate `R` script. The GitHub repository for this work is available here.

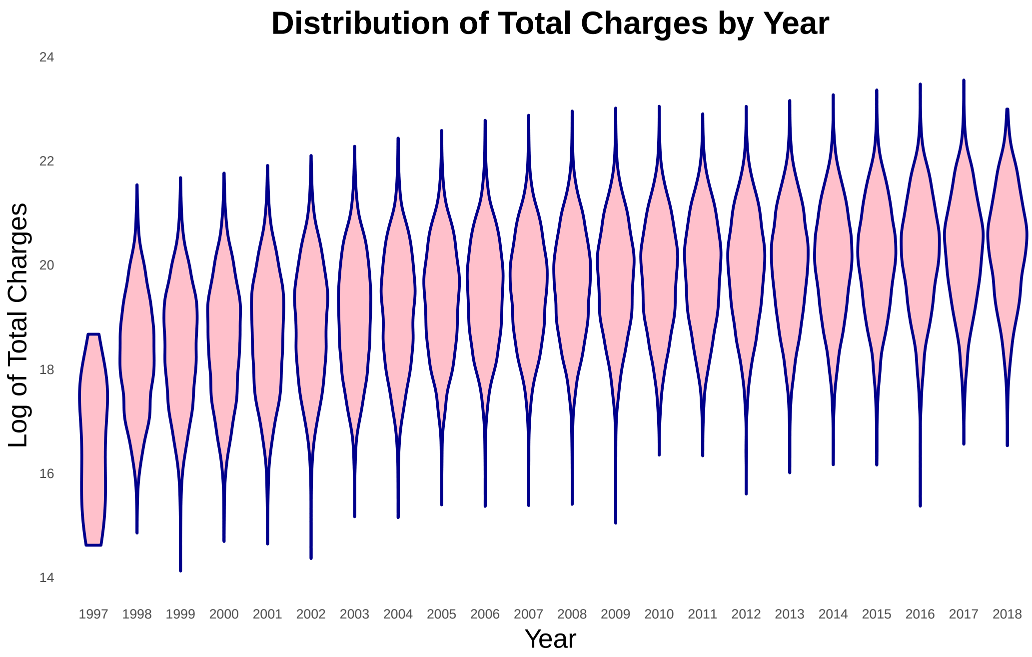
Summarize the Data

Question 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.

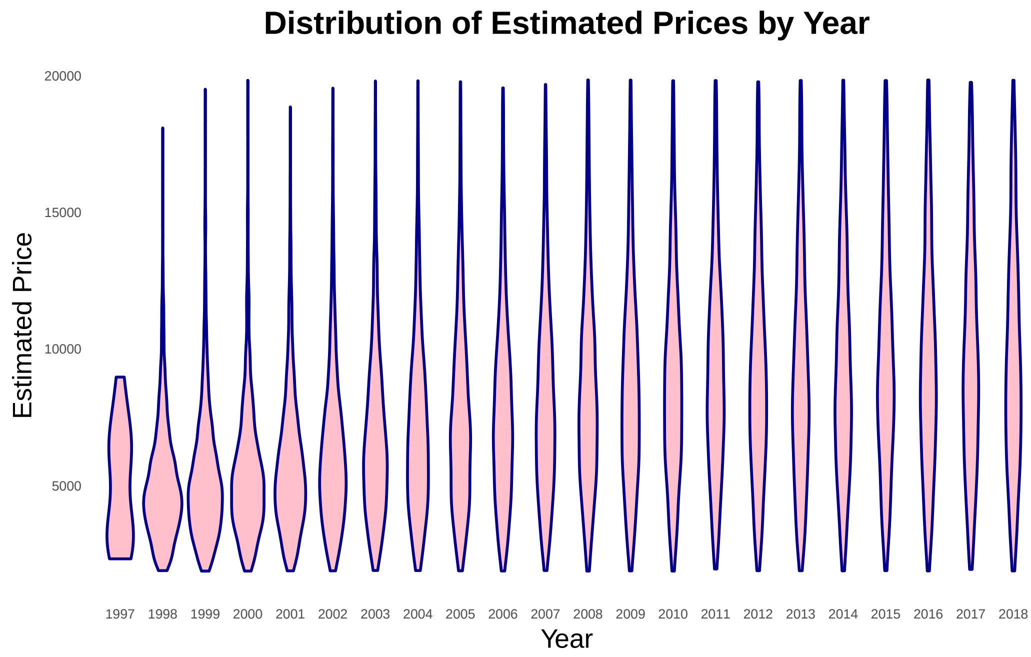


Question 2. After removing/combining multiple reports, how many unique hospital IDs exist in the data? 9323

Question 3. What is the distribution of total charges in each year? Show your results with a "violin" plot.



Question 4. What is the distribution of estimated prices in each year? Present your results with a violin plot.



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

From my analysis, the average price among penalized hospitals is $9657 and $9232 among non-penalized hospitals.

Question 6. Split hospitals into quartiles based on bed size. Provide a table of the average price among treated/control groups for each quartile.

Table 1: Average Prices by Bed Quartile and Penalty

Bed Quartile	Penalty	No Penalty
Q1	7531.701	7675.181
Q2	8260.260	8415.460
Q3	9344.300	9749.992
Q4	11280.274	11561.914

Estimate Average Treatment Effects

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

Table 2: Average Treatment Effect by Estimators

Method	ATE	SE
Nearest Neighbor Matching (IV)	246.3295	171.9265
Nearest Neighbor Matching (Mahalanobis)	246.3295	171.9265
Inverse Propensity Weighting	9225.6907	NA
Simple Linear Regression	NA	339.1507

Question 8. With these different treatment effect estimators, are the results similar, identical, very different? Neareast Neighbor Matching using Inverse Variance distance and Mahalanobis distance yields an identical ATE and SE. It is expected that these estimators would produce consistent results. The Inverse Propensity Weighting method provided a substantially higher ATE. This estimator may suffer from improper weighting or an issue with propensity scores, which is supported by the failure to produce an estimate for SE. The Simple Linear Regression method did not produce an ATE estimate, but the SE suggests that the estimate may have been unreliable. Although results are not complete, the table suggests that different estimators for ATE provide different estimates.

Question 9. Do you think you've estimated a causal effect of the penalty? Why or why not? No, I do not believe the estimates provide a causal effect of the penalty on price. An inherent assumption of causality is that there is no spurious relationships, or that there is no unmeasured confounding variable. The various estimators used in Question 7 cannot control for unmeasured confounders. The NA values also suggest that some models may not be very accurate. This means we cannot rely on these methods to capture the causal effect without bias.

Question 10. Tell me one thing you learned and one thing that really aggravated or surprised you. This dataset was quite large, my computer definitely felt it. It was a bit frustrating working with such a large amount of data but I found it interesting to work with a proper dataset and run code capable at handling all the observations.