

Homework 2

February 17, 2025

Research Methods, Spring 2025

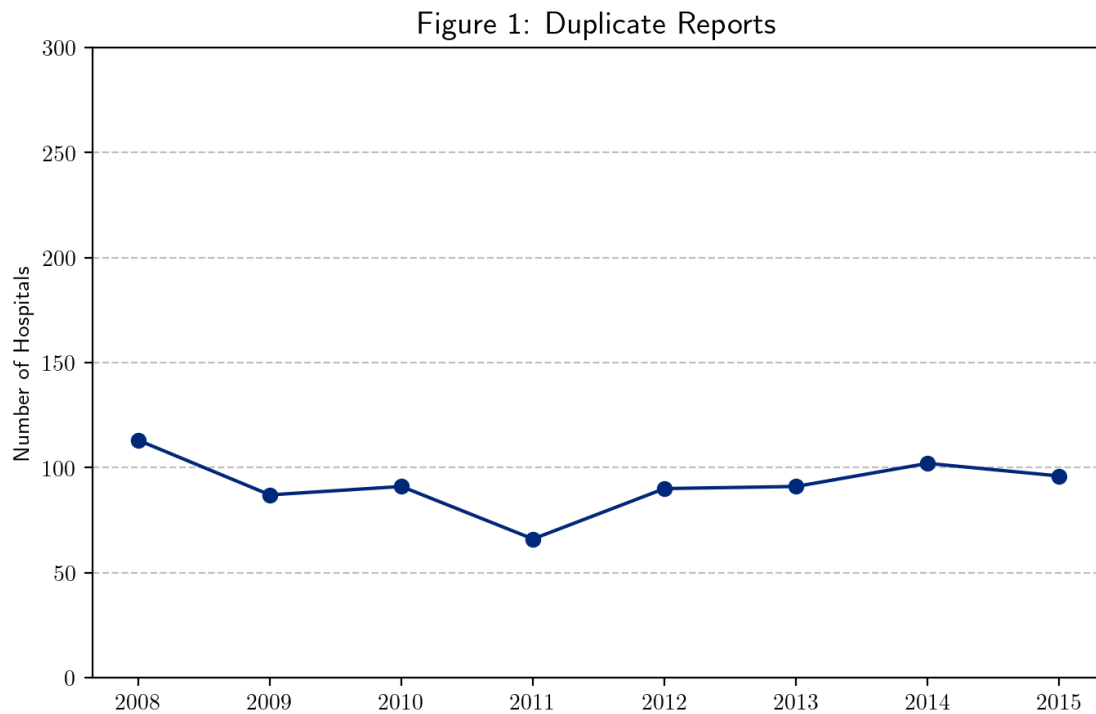
Author: Pablo Estrada

The answers to the homework questions using `Python` are described below. Note that I follow closely the analysis from professor McCarthy and his `R` script. My analysis file is also available in the analysis folder.

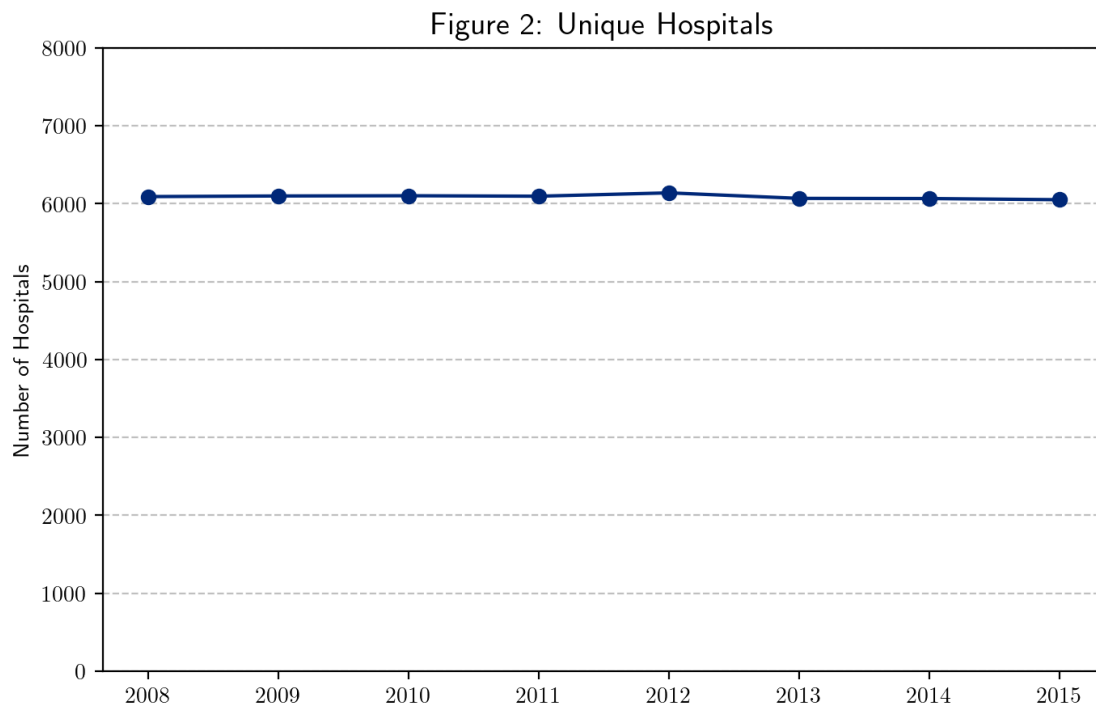
1. How many hospitals filed more than one report in the same year?

Show your answer as a line graph of the number of hospitals over time.

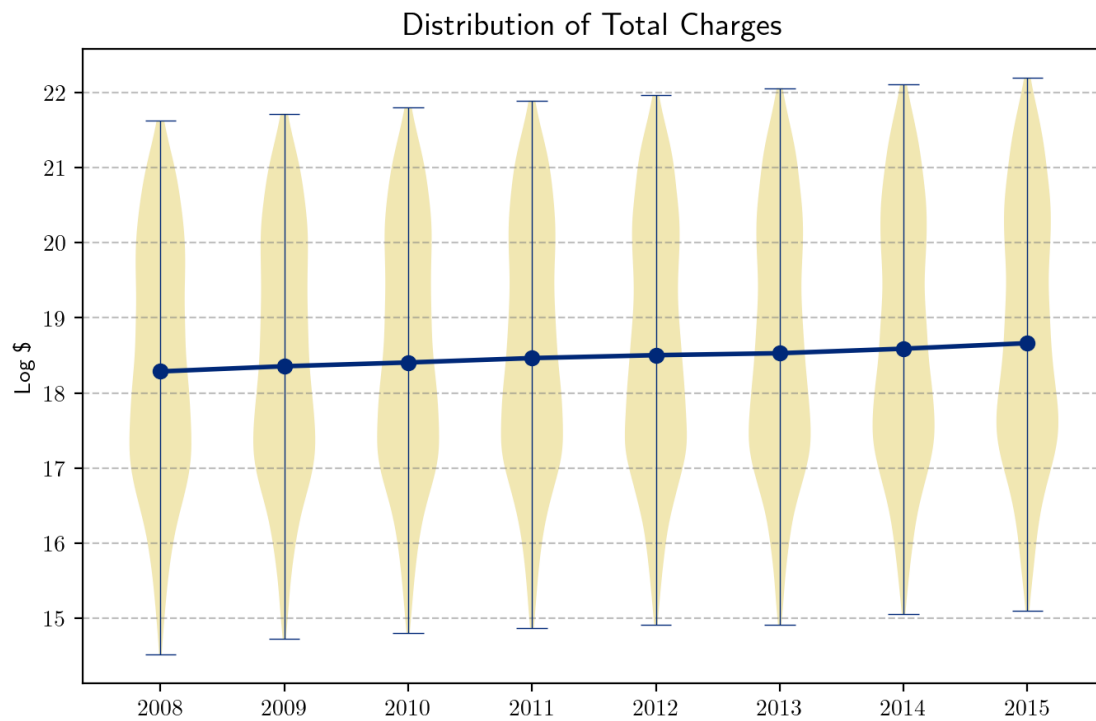
Number of distinct providers: 6746



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

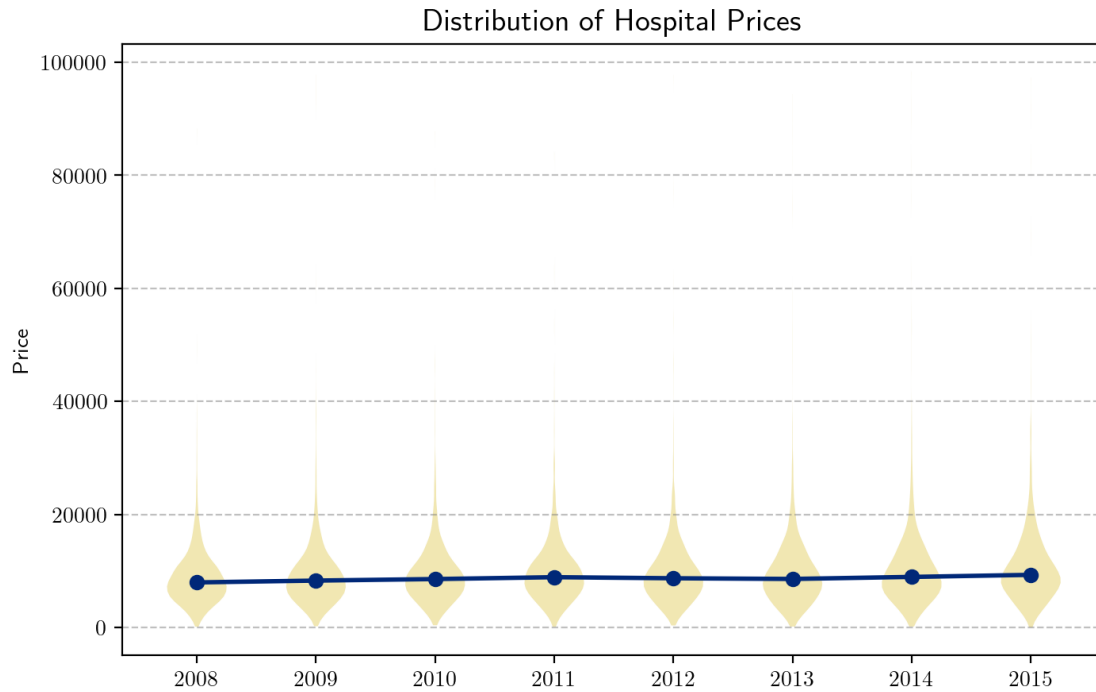


3. What is the distribution of total charges (tot_charges in the data) in each year?
Show your results with a “violin” plot, with charges on the y-axis and years on the x-axis.



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.



For the rest of the assignment, you should include only observations in 2012. So we are now dealing with cross-sectional data in which some hospitals are penalized and some are not.

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

Average price for hospitals with penalty: 9,872.96 and without penalty: 9,597.96

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.

6. Provide a table of the average price among treated/control groups for each quartile.

Bed Size	Not Penalized	Penalized
1	7747.6	8318.4
2	8602.05	8662.35
3	9869.17	10098.1
4	12367.9	12002.1

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

	INV	MAH	IPW	OLS
ATE	124.53	124.53	124.53	124.53
SE	228.63	228.63	234.98	234.89