

# **Homework 2**

## **Submission 1**

Conor Mulligan

This is my first submission of homework 2.

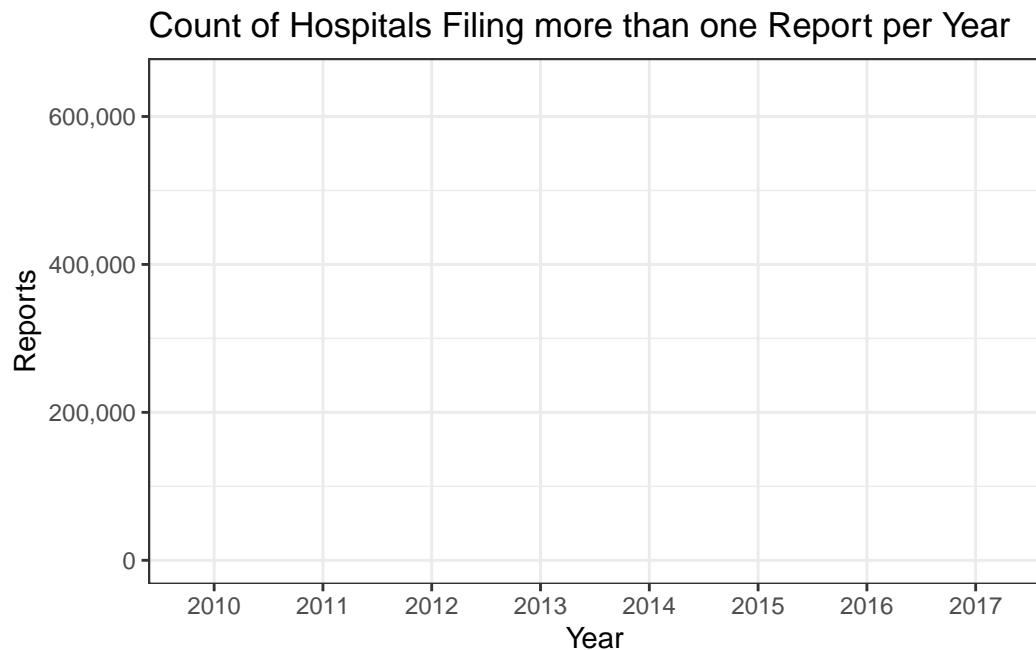
[Link to Github](#)

My answers to the homework questions are below. Analysis is done in a separate R workspace.

## 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.

```
fig.hospital.reports
```



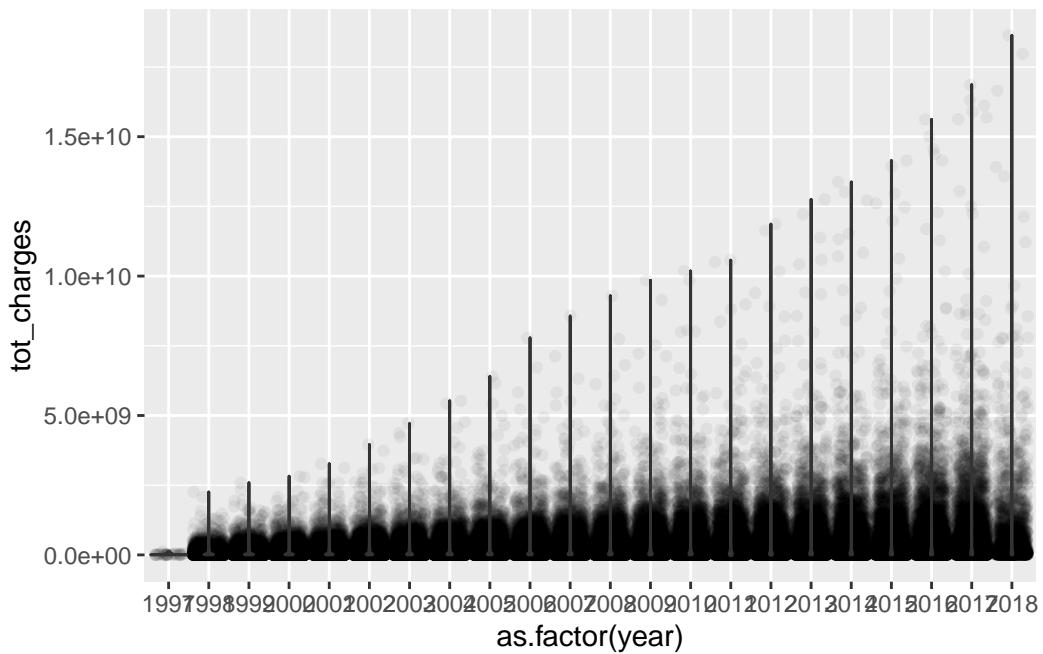
2. After removing/combining multiple reports, how many unique hospital IDs (Medicare provider numbers) exist in the data?

Looking in the data, we are able to see that there are rows, so there are total plan types.

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. .

Warning: Removed 4748 rows containing non-finite values (`stat\_ydensity()`).

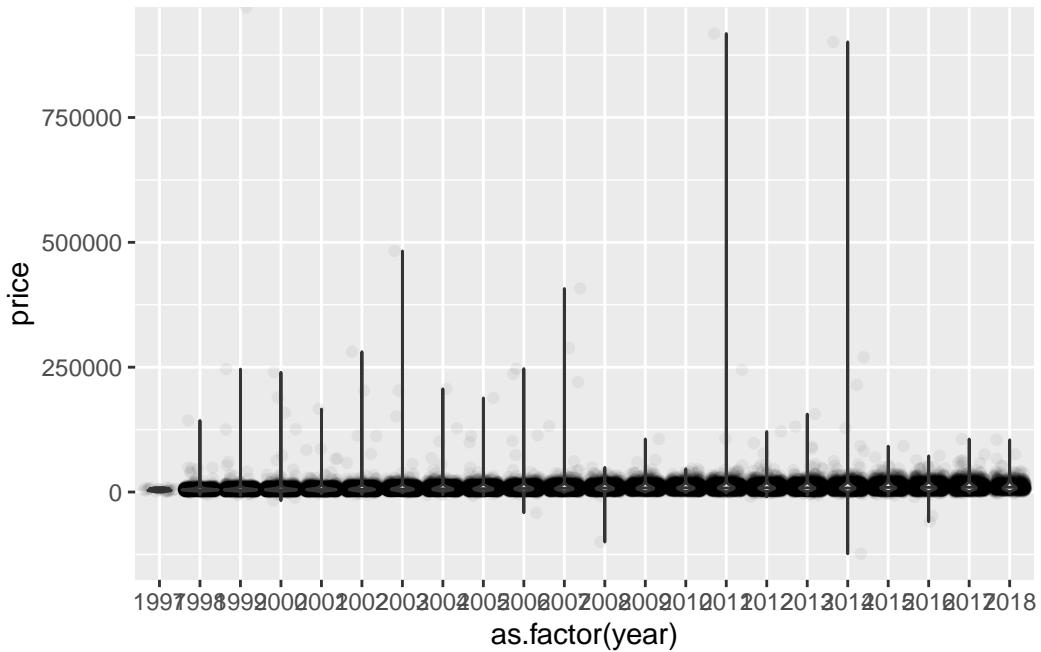
Warning: Removed 4748 rows containing missing values (`geom\_point()`).



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.

Warning: Removed 63663 rows containing non-finite values (`stat\_ydensity()`).

Warning: Removed 63662 rows containing missing values (`geom\_point()`).



## **Estimate ATEs**

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

The average price after calculating between the two groups is 9560 FALSE 9896 TRUE where True is penalized hospitals and False is non penalized hospitals.

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.

My best attempt at the table is below. I was not really sure how to create the quartiles properly and get the hospitals into a binary indicator format.

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

I had some trouble creating the estimators. I was unable to load the matching package, so I could not create them accurately. Due to this I was not really able to also create the table correctly.

## **Summary Questions**

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

As I was unable to accurately source and estimate the treatment indicators, I cannot really tell if the results are similar or different. I would assume the results are similar among the estimators.

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

I think that it is not possible to say that a causal effect was estimated. We might be able to say there is a correlation, but much more analysis based on more data and studies would need to be done to say for certain that the penalty exactly causes an effect.

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 you.

I was unable to properly create the ATEs when looking at the data, which caused some issues. I found it difficult to also create the quartiles and indicator variables for the beds. Overall, I do feel a bit more comfortable with some aspects of the coding process, such as loading everything up and trying to create some visuals.