

Homework 2

Submission 2

Conor Mulligan

This is my first submission of homework 2.

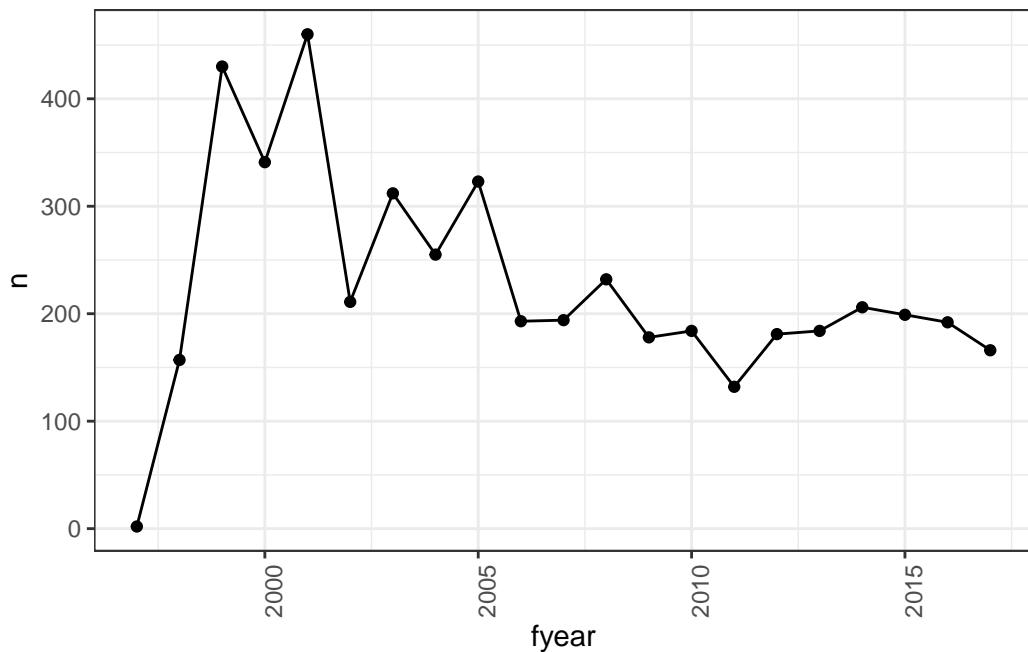
[Link to Github](#)

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
```



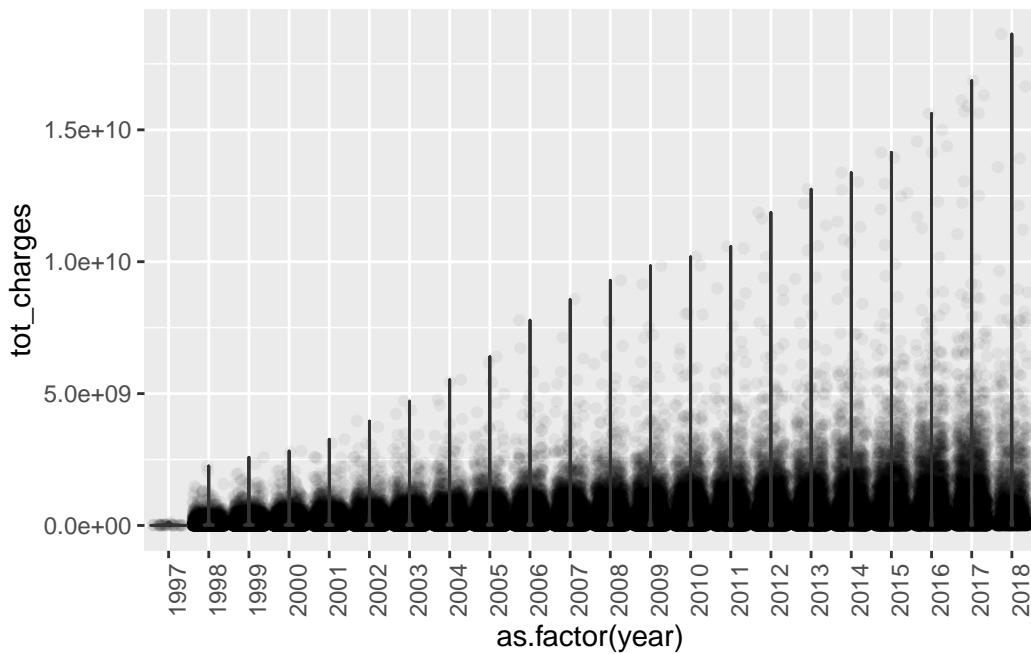
2. After removing/combining multiple reports, how many unique hospital IDs (Medicare provider numbers) exist in the data?

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()`).

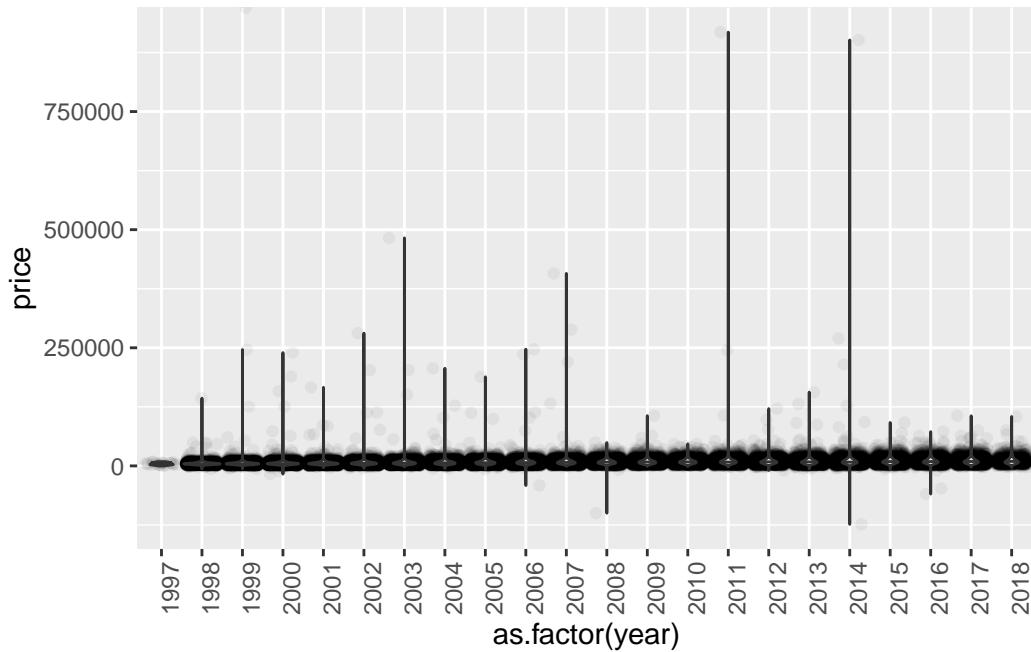
Warning: Removed 4748 rows containing missing values (`geom_point()`).



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 (not sure how to get my answer in the text properly as I did with the unique counts of provider number).

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. As with number 5, false is non-penalized mean price for hospitals and true is penalized mean price for hospitals by quartile.

```
options(knitr.kable.NA = 0)
knitr::kable(table.quartiles,
  col.names=c("Quartiles","Q1","Q2","Q3",
  "Q4"),
  format.args=list(big.mark=","),
  booktabs = TRUE) %>%
  kable_styling(latex_options=c("scale_down"))
```

Warning in styling_latex_scale(out, table_info, "down"): Longtable cannot be resized.

Quartiles	Q1	Q2	Q3	Q4
FALSE	7,696.470	8,525.607	9,848.404	12,367.33
TRUE	8,286.338	8,721.033	10,132.315	12,068.48

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. Specifically, some of the data did not appear to be read properly for the regression, so I could not put that in at present. Further, I was unable to get row names to properly display for the data shown. The first is Mahalanobis, the second is Propensity Score, and the third is Inverse Matching. Regardless of these troubles, I created the table below.

```
options(knitr.kable.NA = 0)
knitr::kable(table.ates,
            col.names=c("Estimated ATEs"),
            format.args=list(big.mark=","),
            booktabs = TRUE) %>%
  kable_styling(latex_options=c("scale_down"))
```

Warning in styling_latex_scale(out, table_info, "down"): Longtable cannot be resized.

Estimated ATEs
193.8313
-201.0278
193.8313

Summary Questions

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

The estimators appear to be quite close to each other, with two of them being the same that I got.

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 could not get some of the data to cooperate as I wanted. The regression analysis did not perform properly despite my best efforts, and is something I will focus on for the final submission. I was able to get the quartiles this time around, so that was a fair improvement. Still, I continue to struggle through and learn as I use the data.