

# **Homework 2**

## **Submission 3**

Conor Mulligan

This is my third 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.

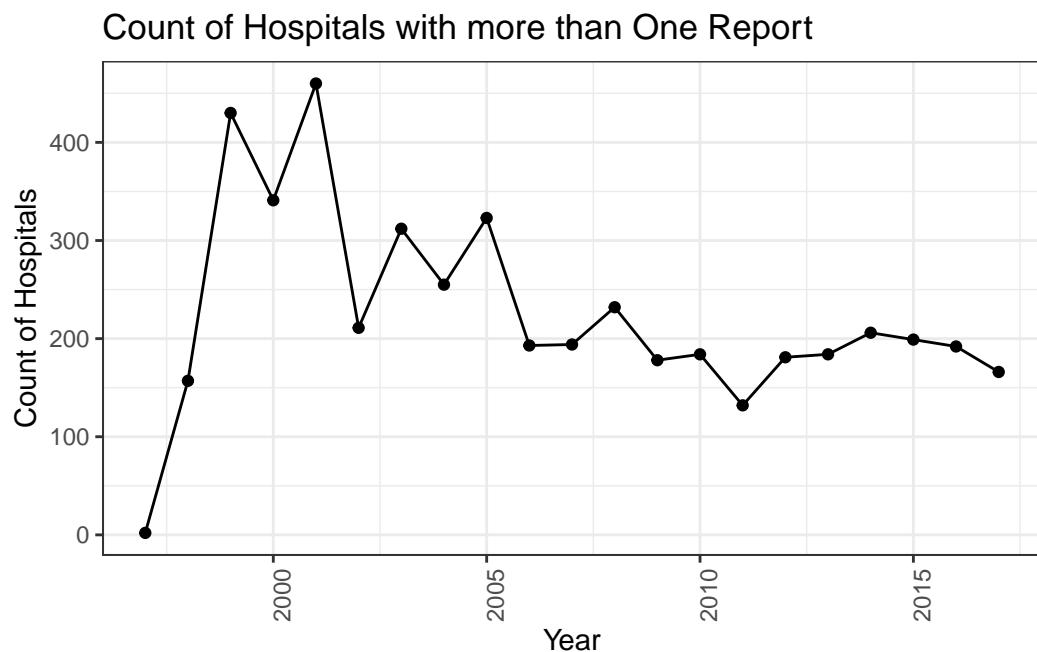
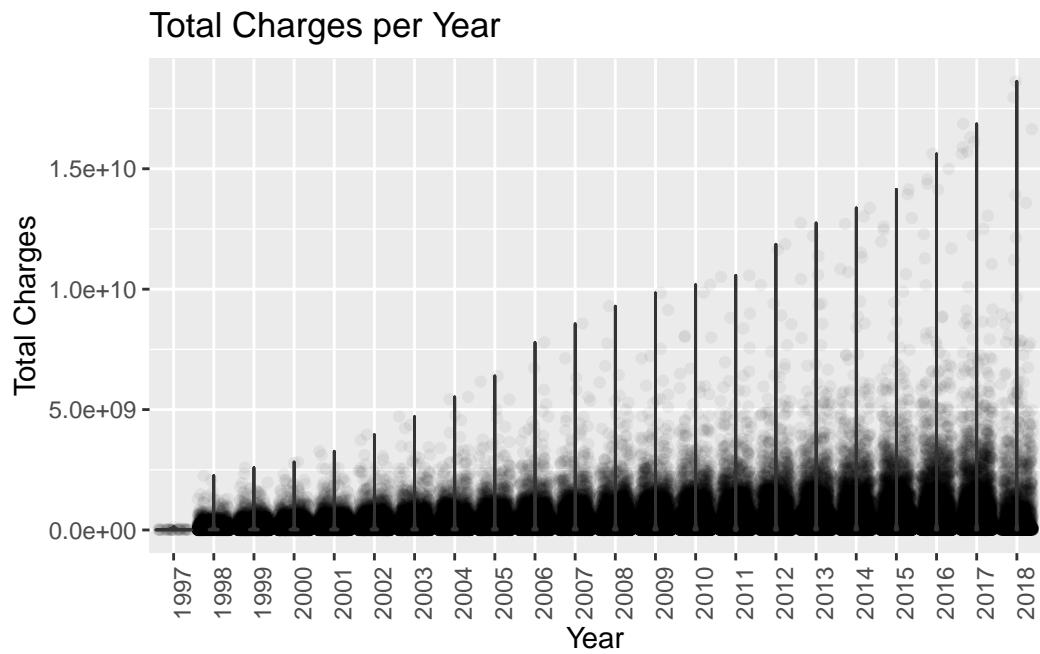


Figure 1: Count of Hospitals with more than one Report over Time

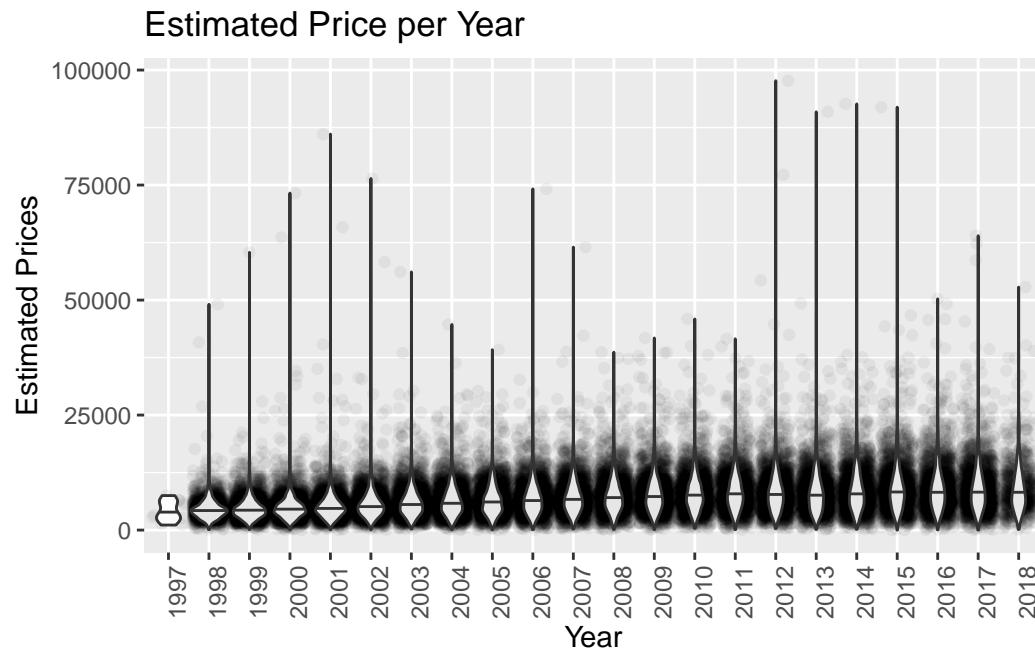
2. After removing/combining multiple reports, how many unique hospital IDs (Medicare provider numbers) exist in the data?

Looking in the data, we are able to see that there are 9323 hospital IDs.

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.



## **Estimate ATEs**

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

The average price after calculating between the two groups is presented in a table below where True is penalized hospitals and False is non penalized hospitals.

Table 1: Average Price of Penalized vs. Non-Penalized Hospitals

Penalty	Average Price
FALSE	9,560.413
TRUE	9,896.308

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.

Table 2: Prices by Bed Size Quartiles and Penalty

Penalty	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.

For the final submission, I was able to get all estimators working properly. However, 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.

Table 3: Estimated ATEs

Estimated ATEs
193.8313
193.8313
193.8313

Below is the results for the regression, which I could not properly include in the table above.

Call:

```
lm(formula = price ~ penalty + first.quartile + second.quartile +
   third.quartile + fourth.quartile + q1.diff + q2.diff + q3.diff +
   q4.diff, data = reg.dat)
```

Residuals:

Min	1Q	Median	3Q	Max
-11592	-3285	-970	2083	89992

Coefficients: (2 not defined because of singularities)

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	12367.3	243.7	50.754	< 2e-16 ***
penaltyTRUE	193.8	239.9	0.808	0.419
first.quartile	-4670.9	337.1	-13.855	< 2e-16 ***
second.quartile	-3841.7	343.9	-11.170	< 2e-16 ***
third.quartile	-2518.9	348.5	-7.227	6.39e-13 ***
fourth.quartile	NA	NA	NA	NA
q1.diff	888.7	696.6	1.276	0.202
q2.diff	494.3	669.5	0.738	0.460
q3.diff	582.8	658.2	0.885	0.376
q4.diff	NA	NA	NA	NA

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5449 on 2725 degrees of freedom  
Multiple R-squared: 0.08959, Adjusted R-squared: 0.08725  
F-statistic: 38.31 on 7 and 2725 DF, p-value: < 2.2e-16

## **Summary Questions**

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

The estimators appear to be identical to each other, with all of them being 193.8 after estimating based on quartiles for bed size.

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 was able to run properly this time, but I could not properly include the results in one table alongside the other ATEs. Other than that, I tried to improve the clarity of my graphs and the readability of my overall final document, which I think I mostly succeeded in doing.