Homework 2

Research in Health Economics, Spring 2024
https://github.com/riadharnidharka/Econ470hw2

Ria Dharnidharka

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.

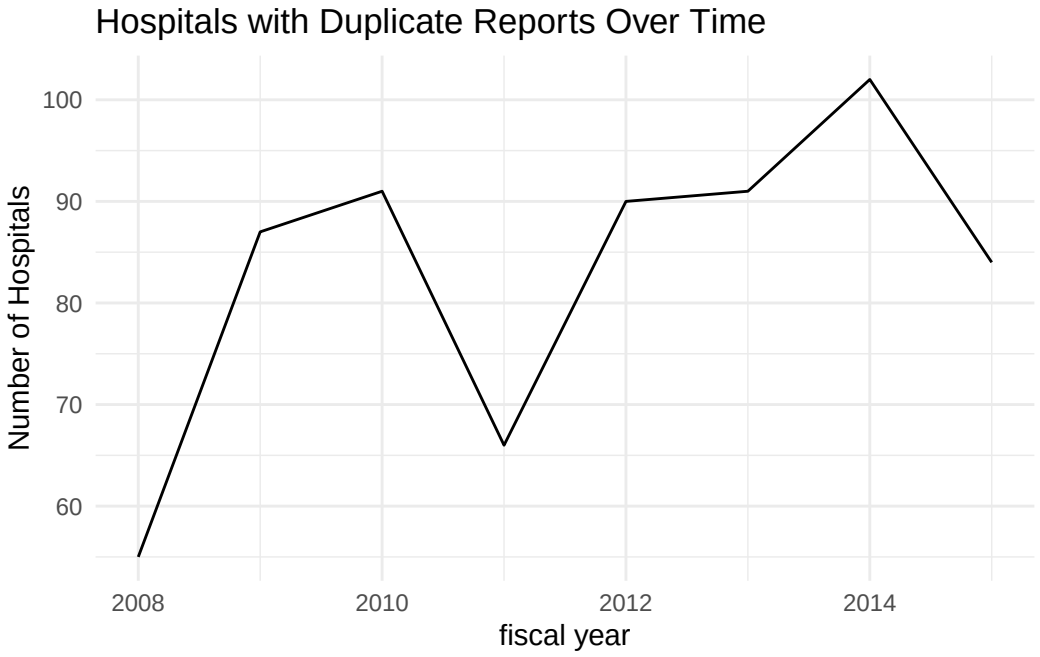

Figure 1: Hospitals with Duplicates

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

There are 48803 unique hospital ids in the dataset

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. For a nice tutorial on violin plots, look at Violin Plots with ggplot2.

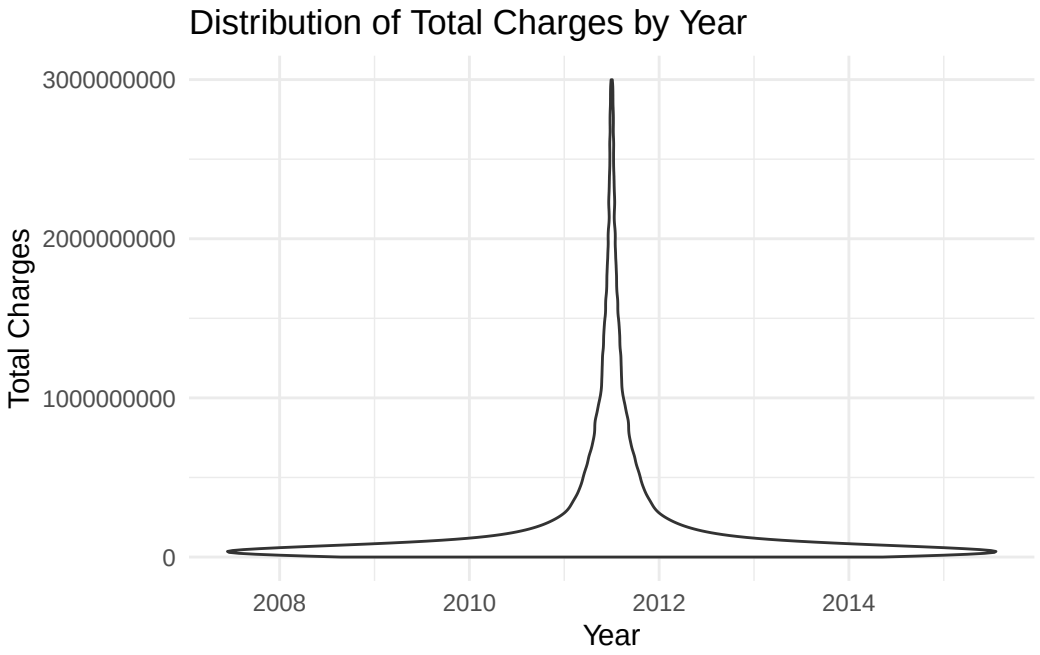

Figure 2: Total Charges

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.

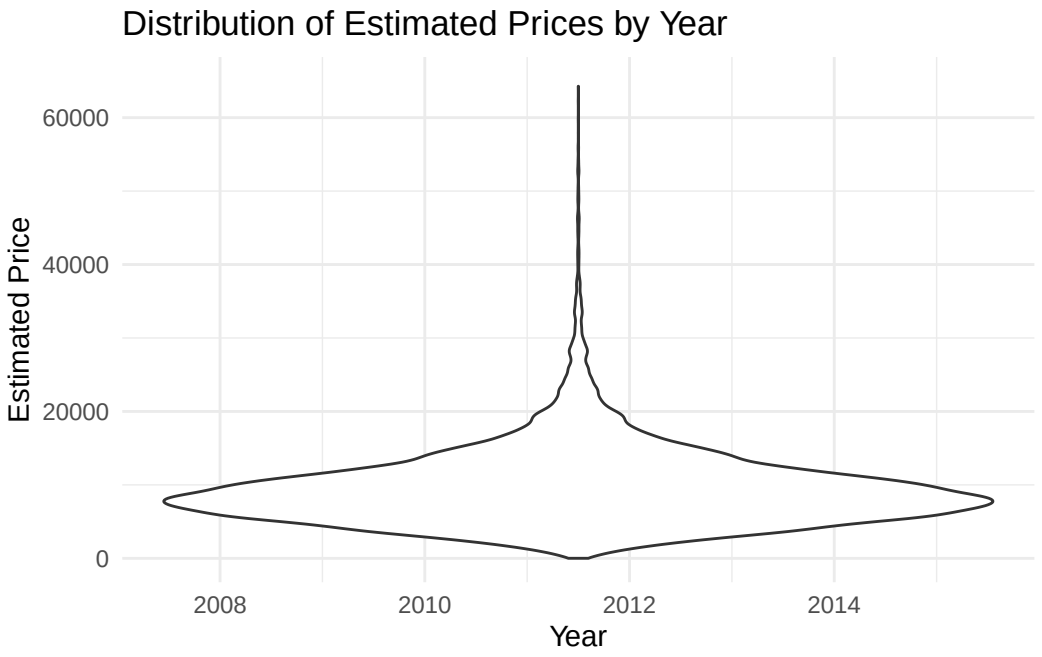

Figure 3: Estimated Prices

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

The average price among non-penalized hospitals is $9560.41, while the average price among penalized hospitals is $9896.31.

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.

Table 1: Quartile Price Summary

Quartile	Treated Avg Price	Non-Treated Avg Price
1	8,318.709	7,684.240
2	8,690.891	8,510.959
3	10,127.130	9,856.928
4	12,068.479	12,355.606

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

I attempted this question using the following code, but could not figure out what to replace these variables with in order to use our data. For example, I could not figure out what to put in for the treatment and outcomes. I will continue working on this question for the second submission.

#Perform nearest neighbor matching match_iv <- matchit(treatment ~ beds, data = your_data_frame, method = "nearest", distance = "genetic")

#Calculate Mahalanobis distance matrix mah_dist <- mahalanobis(your_data_frame[, c("bed_size_variable", "other_covariates")])

#Perform nearest neighbor matching (1-to-1) with Mahalanobis distance match_mah <- Match(your_data_frame$treatment, your_data_frame$bed_size_variable, Mahalanobis = TRUE)

#Calculate inverse propensity scores model_ps <- glm(treatment ~ bed_size_variable, data = your_data_frame, family = binomial) propensity_scores <- predict(model_ps, type = "response")

#Perform inverse propensity weighting weighted_data <- your_data_frame weighted_data$weights <- ifelse(weighted_data$treatment == 1, 1 / propensity_scores, 1 / (1 - propensity_scores))

#Perform simple linear regression model_lm <- lm(outcome ~ factor(bed_size_variable) + factor(treatment) + factor(bed_size_variable) * factor(treatment), data = your_data_frame) summary(model_lm)

#Obtain robust standard errors for the regression coefficients coeftest(model_lm, vcov = sandwich)

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

I won't be able to fully answer this question until I have figured out question 8, but I would imagine that the results would be similar with the different treatment effects. I would guess this because using slightly different methods would mean there could not be identical results, but estimates should not be so different because they are al aiming to estimate the same thing.

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

Once again I cannot fully answer the question until I have answered question 8 to see the results, but I would guess that we are estimating a causal effect of the penalty. I would guess this because when we use weighting in our matching, we are able to control more for different factors that would interfere with causality.

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 or surprised you.

My experience working with the data this time around was much better, and I felt a lot more comfortable every step of the way. One thing I learned is how to do the analysis of the data without looking at the exact code, and instead using the slides and chatgbt to guide me in creating my own code. One thing that aggravated me is not knowing how to manipulate standard lines of code to incorprate our data, such as in question 7.