Homework 4

ECON 470, Spring 2025

Ethan Murakami

Here is a link to my repository: {https://github.com/bemur3/hwk4}

1. Remove all SNPs, 800-series plans, and prescription drug only plans (i.e., plans that do not offer Part C benefits). Provide a box and whisker plot showing the distribution of plan counts by county over time. Do you think that the number of plans is sufficient, too few, or too many?

```
Warning: Removed 696 rows containing non-finite outside the scale range
(`stat_boxplot()`).
```
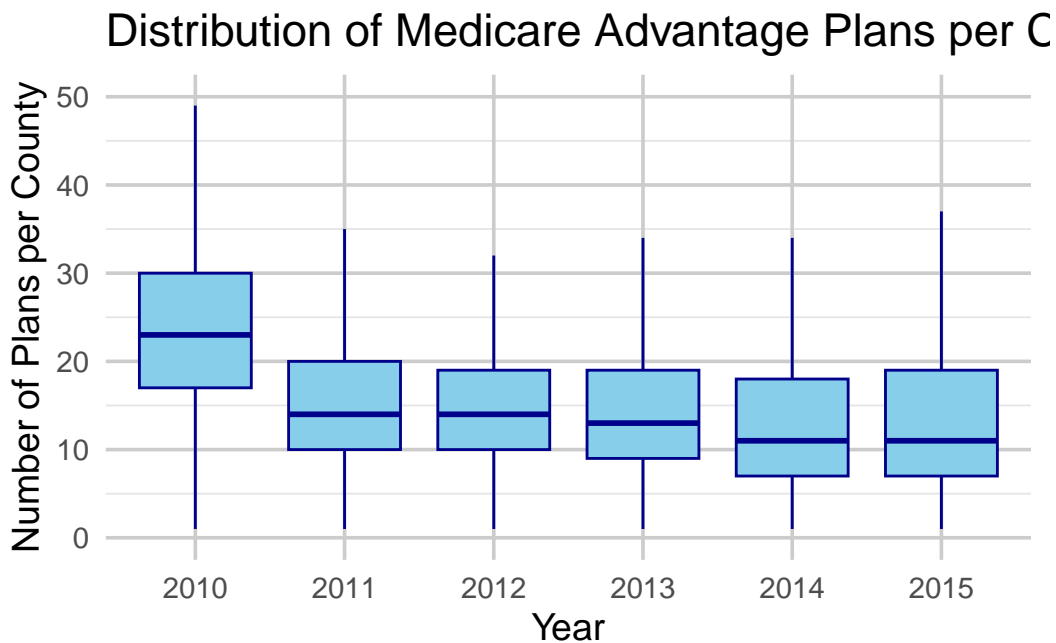

The distributions of plan counts seem to slightly decrease each year. Aside from 2010 though, the median number of plans appears to stay consistent. The range of 10-20 plans is sufficient, as it allows beneficiaries variation between different premiums, benefits and quality while also not being too overwhelming in number.

2. Provide bar graphs showing the distribution of star ratings in 2010, 2012, and 2015. How has this distribution changed over time?

MA Part C Star Rating Distributions (2010, 2(
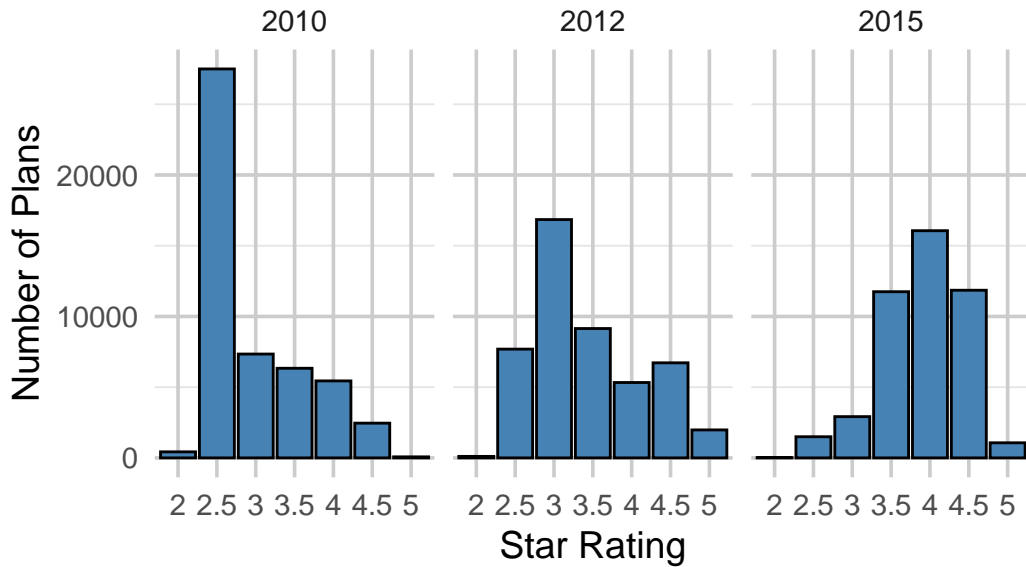
The distribution shifts to be skewed more right towards the higher star ratings over time, with 2.5 star ratings in particular having drastic decreases. This suggests a positive trend towards higher quality plans.

3. Plot the average benchmark payment over time from 2010 through 2015. How much has the average benchmark payment risen over the years?

Average MA Benchmark Payment (2010–2015)
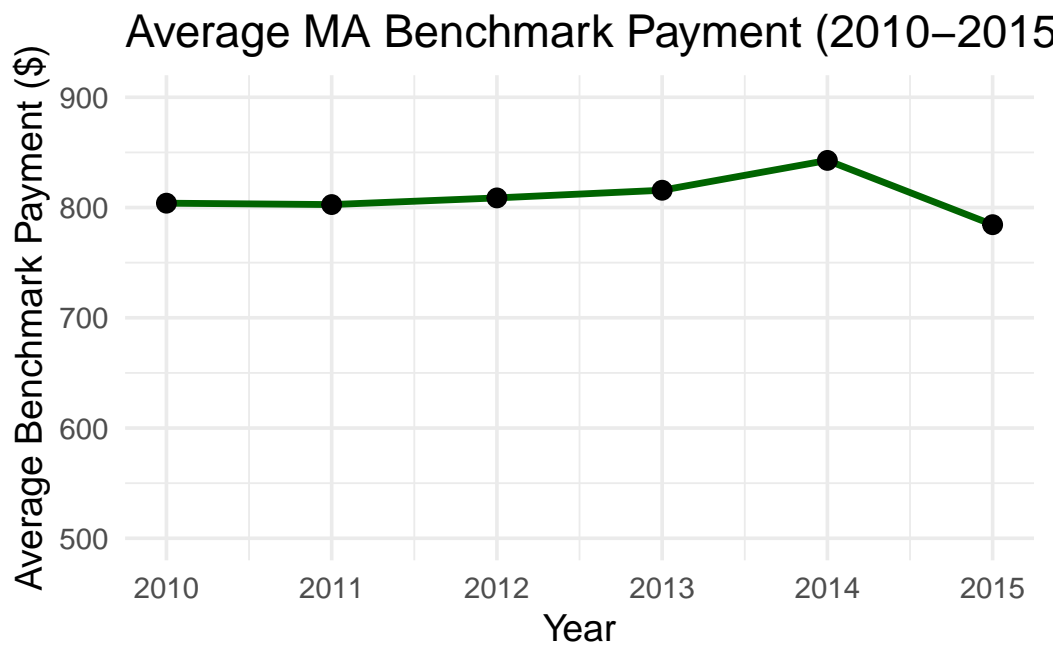
Average benchmark payment steadily rose from 2010-2014 before dropping by about $800 in 2015.

4. Plot the average share of Medicare Advantage (relative to all Medicare eligibles) over time from 2010 through 2015. Has Medicare Advantage increased or decreased in popularity? How does this share correlate with benchmark payments?

Average Share of MA Enrollment (2010–2015)
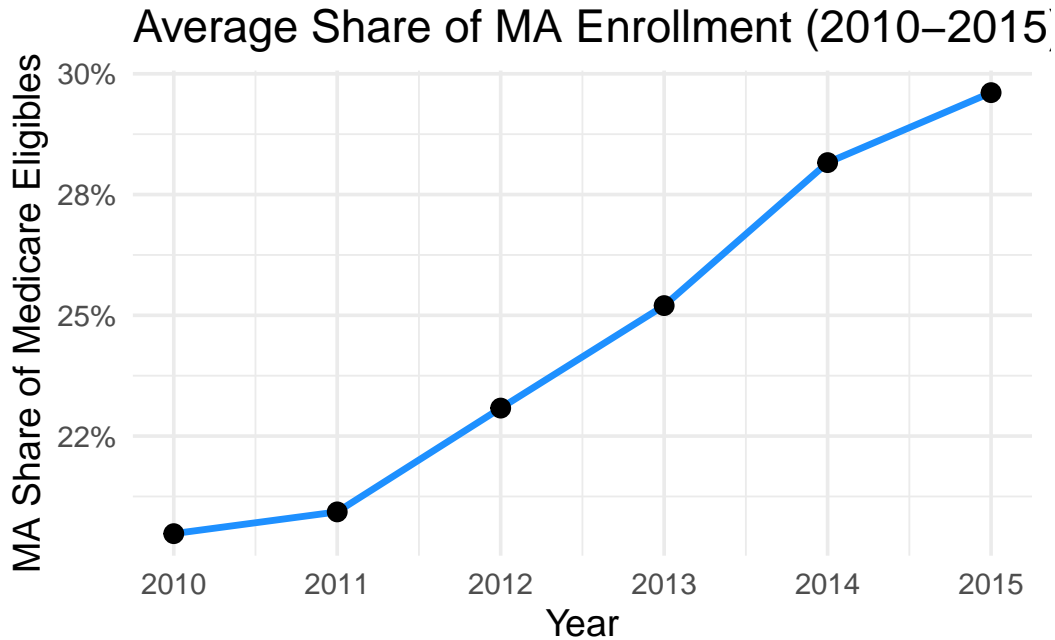
The enrollments steadily increase each year, showing a growing popularity in Medicare Advantage. As benchmark payments increased or remained relatively elevated, they likely facilitated the expansion of MA plans, contributing to the rise in MA enrollment share. This effect seems to be independent of benchmark payments, as they stayed mostly consistent in question 3.

5. Calculate the running variable underlying the star rating. Provide a table showing the number of plans that are rounded up into a 3-star, 3.5-star, 4-star, 4.5-star, and 5-star rating.

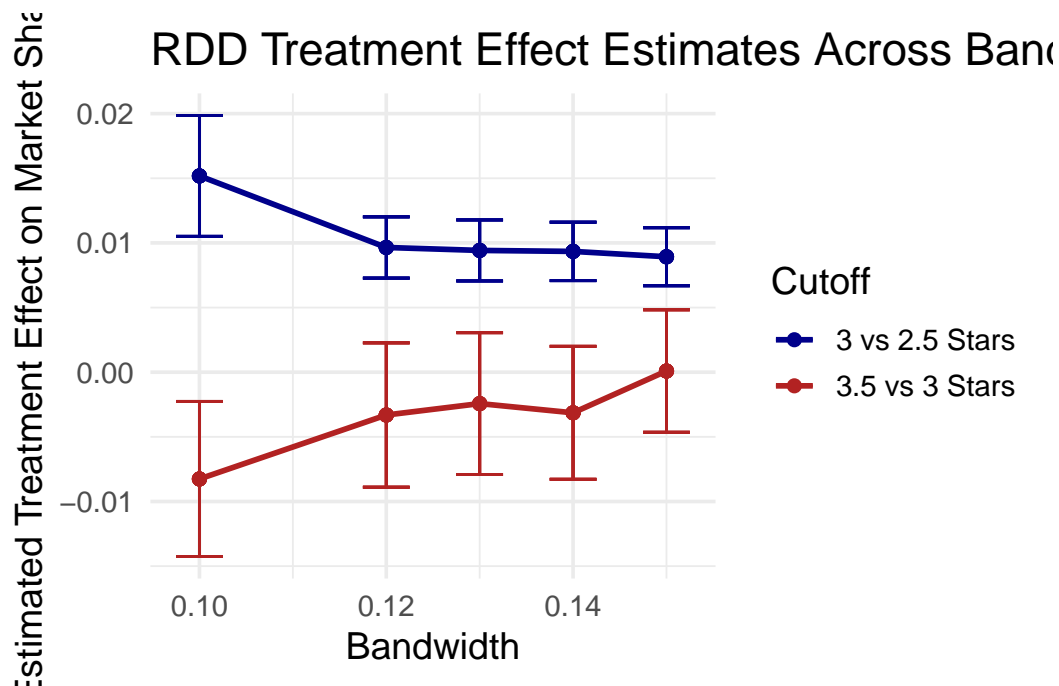Table 1: Number of Plans Rounded Up to Each Start Ratint(2010)

Star_Rating	rounded.up
3.0	1346
3.5	1555
4.0	1402
4.5	176
5.0	26

6. Using the RD estimator with a bandwidth of 0.125, provide an estimate of the effect of receiving a 3-star versus a 2.5 star rating on enrollments. Repeat the exercise to estimate the effects at 3.5 stars, and summarize your results in a table.

Table 2: RDD Regression Summary (Bandwidth = 0.125)

| Cutoff	Intercept	Rounded Up	Running Score	Num Obs	R-squared	RMSE
3 vs 2.5 Stars	0.009	0.010	-0.024	4483	0.019	0.015
3.5 vs 3 Stars	0.017	-0.003	0.022	1663	0.001	0.026

7. Repeat your results for bandwidhts of 0.1, 0.12, 0.13, 0.14, and 0.15 (again for 3 and 3.5 stars). Show all of the results in a graph. How sensitive are your findings to the choice of bandwidth?

RDD Treatment Effect Estimates Across Ban

Estimated Treatment Effect on Market Sha

Cutoff
— 3 vs 2.5 Stars
— 3.5 vs 3 Stars

Bandwidth

The results for the 3 star cutoff were mostly insensitive to bandiwdth. The results for the 3.5 star cutoff were slightly more sensitive between bandwidths, as the treatment effect jumped from a negative value to around 0 from 0.14 to 0.15. Overall though, the results are largely insensitive, demonstrating confidence in the RD estimates.

8. Examine (graphically) whether contracts appear to manipulate the running variable. In other words, look at the distribution of the running variable before and after the relevent threshold values. What do you find?

Density of the Running Variable Near RD Thresholds

(a) Around 3.0 Cutoff   (b) Around 3.5 Cutoff
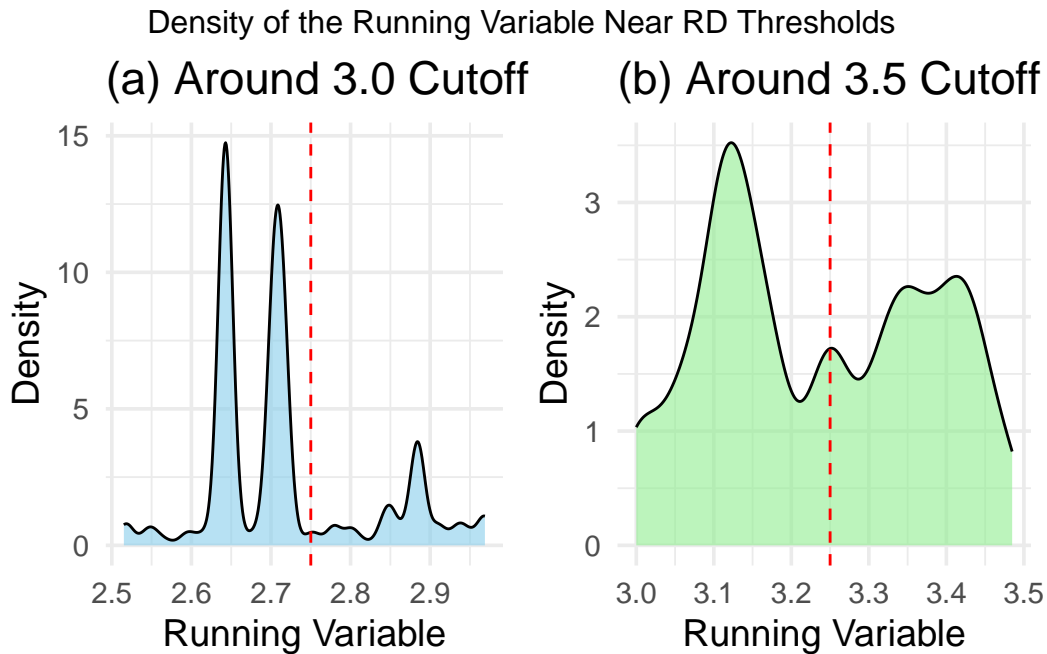
There is no clear evidence of insurer manipulation here. If there was, the distribution of plans would be skewed towards above the cutoff threshold because insurers would want to push their scores just slightly above the threshold, but the distribution for both sides is skewed towards the left of the threshold. It could be argued that there may have been some attempt at manipulation for the 3.0 cutoff graph, as there is a large spike just before the threshold line, but it ultimately doesn't cross over 2.75.

9. Similar to question 4, examine whether plans just above the threshold values have different characteristics than contracts just below the threshold values. Use HMO and Part D status as your plan characteristics.
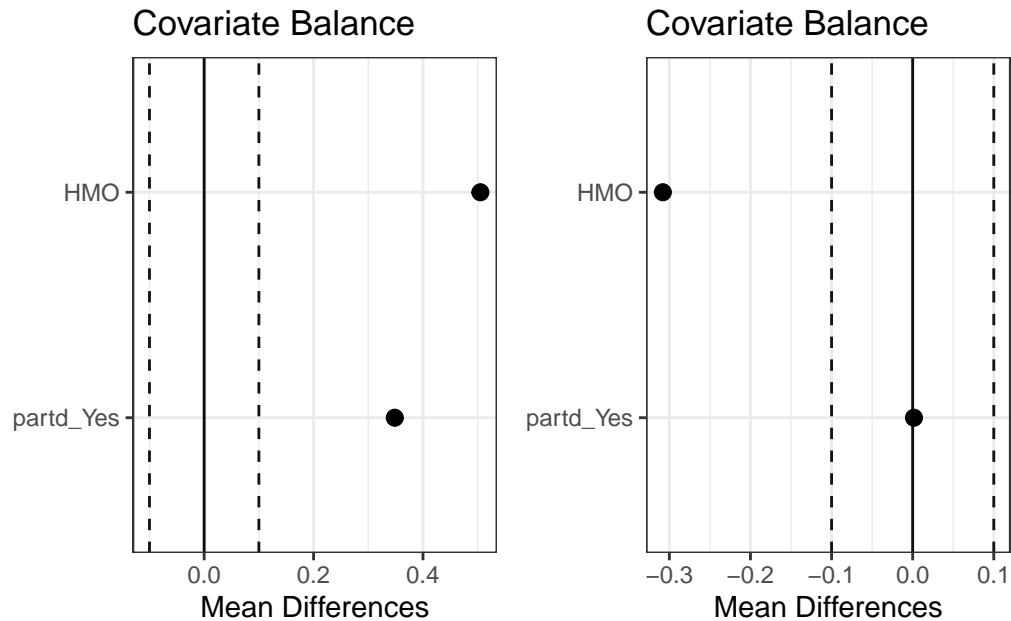
In both plots, we observe that the HMO characteristic exhibits a greater standardized mean difference around the cutoff compared to the Part D status. This indicates that plans just above the rating thresholds (3 and 3.5 stars) are more likely to differ in their HMO status than their Part D offering, suggesting systematic differences between treated and control groups. These imbalances challenge a key assumption of the regression discontinuity design — that units on either side of the cutoff are comparable — and may weaken the internal validity of the RDD estimates.

10.Summarize your findings from 5-9. What is the effect of increasing a star rating on enrollments? Briefly explain your results.

The results of questions 5-9 suggest that increasing star rating can have a significant impact on plan enrollment. Receiving a 3-star rating (vs. 2.5) caused a statistically significant increase in enrollment share. This implies that beneficiaries were more likely to enroll in plans that just crossed the threshold into a 3-star rating, highlighting the salience of star ratings in consumer choice. Surprisingly, plans receiving a 3.5-star rating (vs. 3) saw a decrease in market share. This could potentially mean enrollee preferences don't put as much weight on star ratings within the same number range; a 0.5 jump doesn't have the same impact as jumping up to a full 3 stars as it does jumping up to a 3.5 stars. Question 7 suggests these results were robust across different bandwidths, which supports the validity of our findings, as treatment effects were not sensitive to these bandwidths. Question 9 showed that there is evidence of imbalance at the 3.0-star cutoff, especially for HMO status, which challenges the validity of RDD assumptions there. In contrast, the 3.5-star cutoff appears more suitable for RDD, as observed covariates are better balanced across the threshold. Applying this to our question 8 results, our notion that there was no manipulation becomes even stronger. Despite this, 3.0 stars is a meaningful threshold, associated with measurable changes in market behavior, meaning that raising it above 3 stars could significantly impact its competitiveness.