Homework 4

Research Methods, Spring 2024

Answer Key

My answers to the homework questions are described below. As with the previous homework assignments, note that my analysis is in a seperate `R` script. My analysis file is available here.

Summarize the data

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?

The goal of this graph is to illustrate how many plans are available to an average enrollee in an average county. We remove SNPs, 800-series plans, and Part D only plans in order to focus on a more similar product that is available to everyone. For consistency with the rest of the analysis, I've also removed plans with missing plan IDs, plans in US territories, and plans not operating in an approved service area. The resulting box and whisker plot is provided in Figure 1, which provides a sense of the distribution of plan counts across counties in each year.
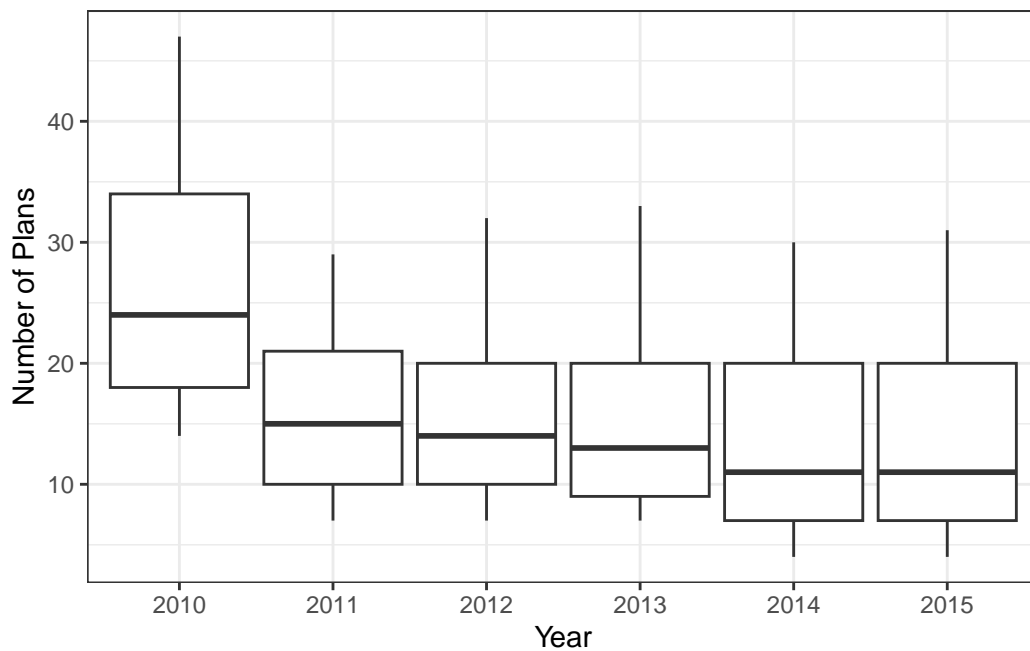
Figure 1: Plan Counts by County over Time

2. Provide bar graphs showing the distribution of star ratings in 2009, 2012, and 2015. How has this distribution changed over time?

Counts of plans by star rating and year are provided in Figure 2. As is evident from the figure, the distribution of ratings has shifted rightward over time, with many more plans receiving higher ratings.
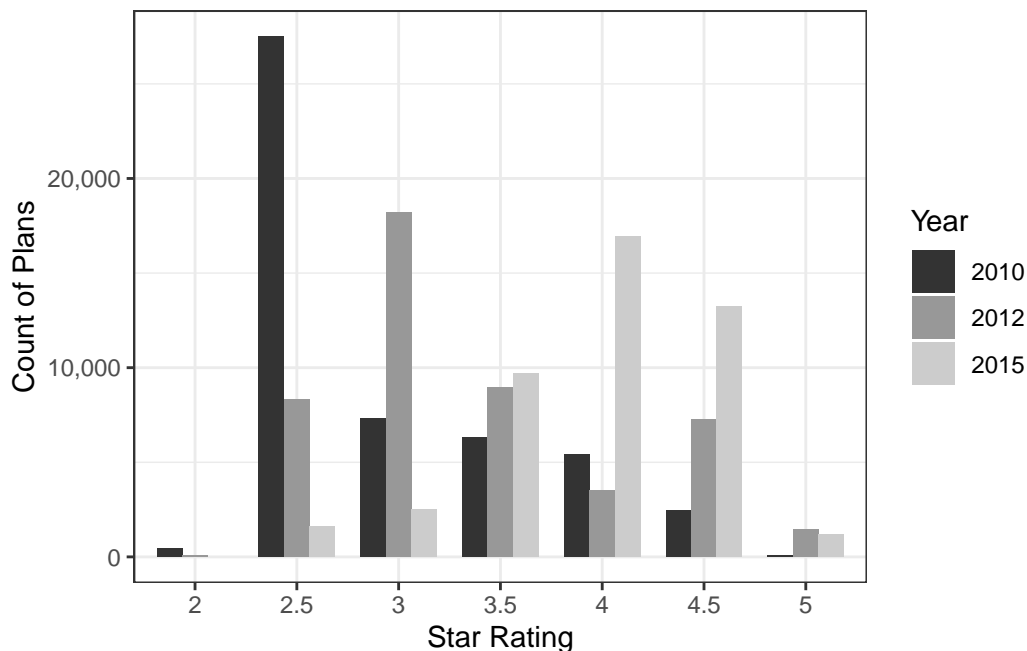
Figure 2: Star Ratings by Year

3. Plot the average benchmark payment over time. How much has the average benchmark payment risen over the years?

Average MA benchmark rates are presented in Figure 3. As we can see, the payment rates have been relatively steady over this time period, with a spike in 2014 and a drop in 2015. Note that 2012 through 2014 is when there were additional quality improvement incentives built into the benchmark payments. Those incentives changed in 2015, consistent with the growth in benchmark payments through 2014 and subsequent decrease.
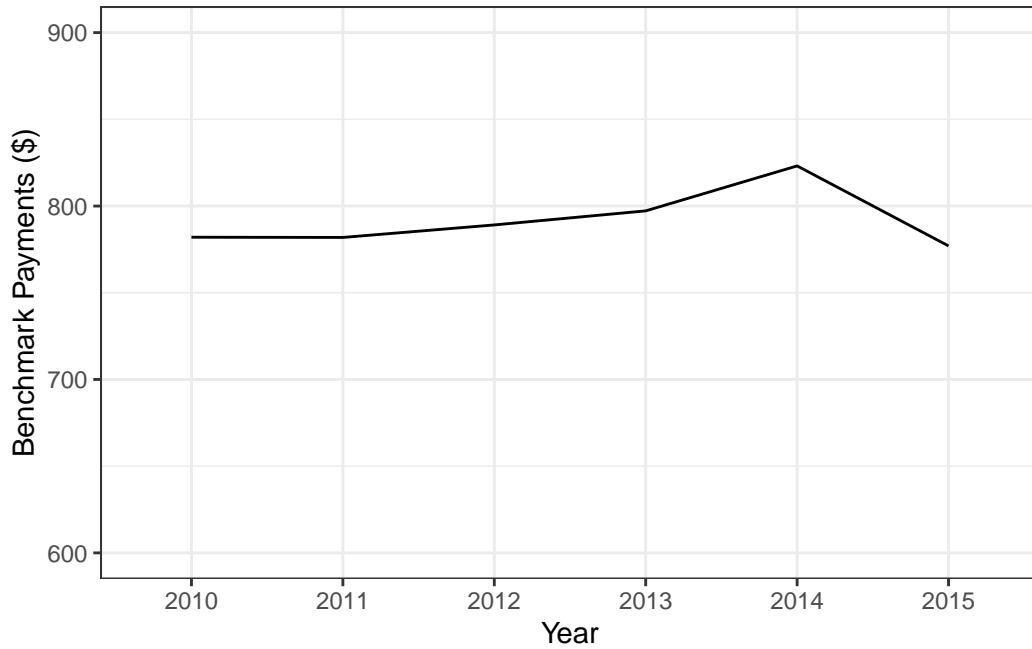
Figure 3: Average Benchmark Payments over Time

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

Average MA market shares are presented in Figure 4, revealing the steep increase in Medicare Advantage popularity over time. This doesn't seem to be strongly related to Medicare Advantage benchmark payments, however, as the correlation between average benchmark payments and market shares is just 0.0737. So the relationship between these two variables is relatively weak.
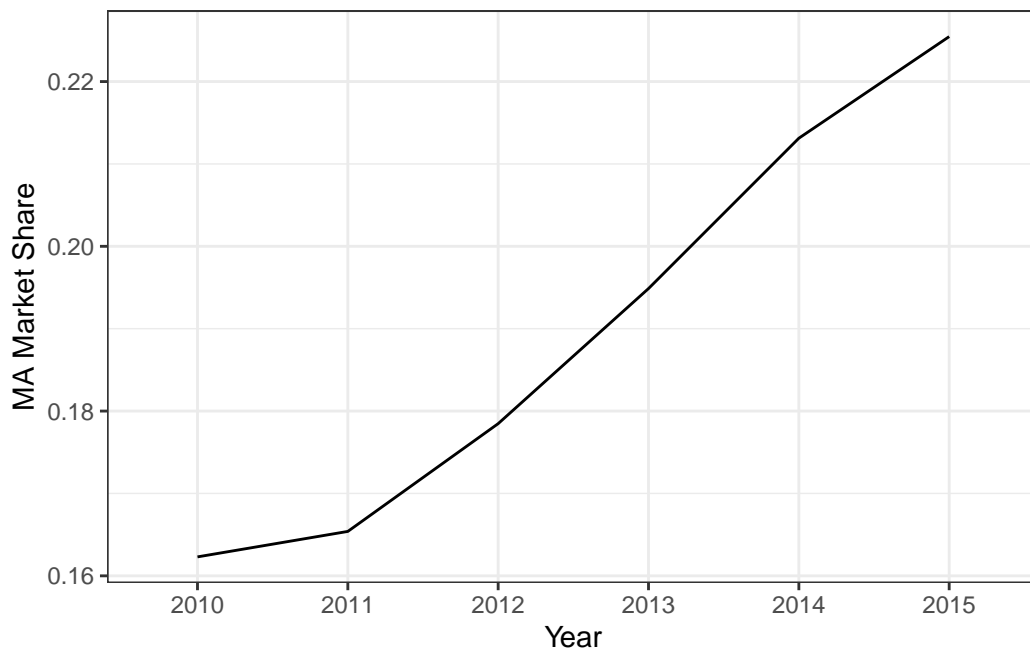
Figure 4: Medicare Advantage Market Share over Time

Estimate ATEs

Now let's work on estimating the effects of quality ratings on enrollments using a regression discontinuity design. We'll focus only on 2009, as this is the year in which the star rating running variable is easiest to replicate.

1. 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.

Note that there are no 5-star plans in 2009, so that's irrelevant in this case. Table 1 presents the counts of plans rounded up into each relevant star rating category.

Table 1: Count of Rounded Plans by Star Rating

Star Rating	Rounded
3.0	1734
3.5	1815
4.0	606
4.5	0
5.0	0

2. 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, 4 stars, and 4.5 stars. Summarize your results in a table.

We know from Table 1 that there are no plans rounded into the 4.5 star category, so we don't need to worry about that one. For the other categories, we just need to run a local linear regression within the relevant window around each star rating. Results are summarized in Table 2.

Table 2: RD Estimates by Star Ratings

	3 Stars	3.5 Stars
(Intercept)	0.010	0.014
	(0.000)	(0.002)
Rounded	0.005	0.003
	(0.001)	(0.002)
Running Score	−0.012	−0.026
	(0.003)	(0.009)
Num.Obs.	5084	1733
R2	0.017	0.005
RMSE	0.02	0.03

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

The hardest part about this question is presenting the answers in a graph. There are **lots** of ways to do this in R (as is the case for most things in R). But the easiest way I've found is to use the `broom` package, which allows you to put other `R` objects (like regression results) into a tidy data set, which you can then plot using standard methods. Figure 5 presents the different RD estimates (from a local linear regression with constant slope terms) across star rating categories and different bandwidth values.
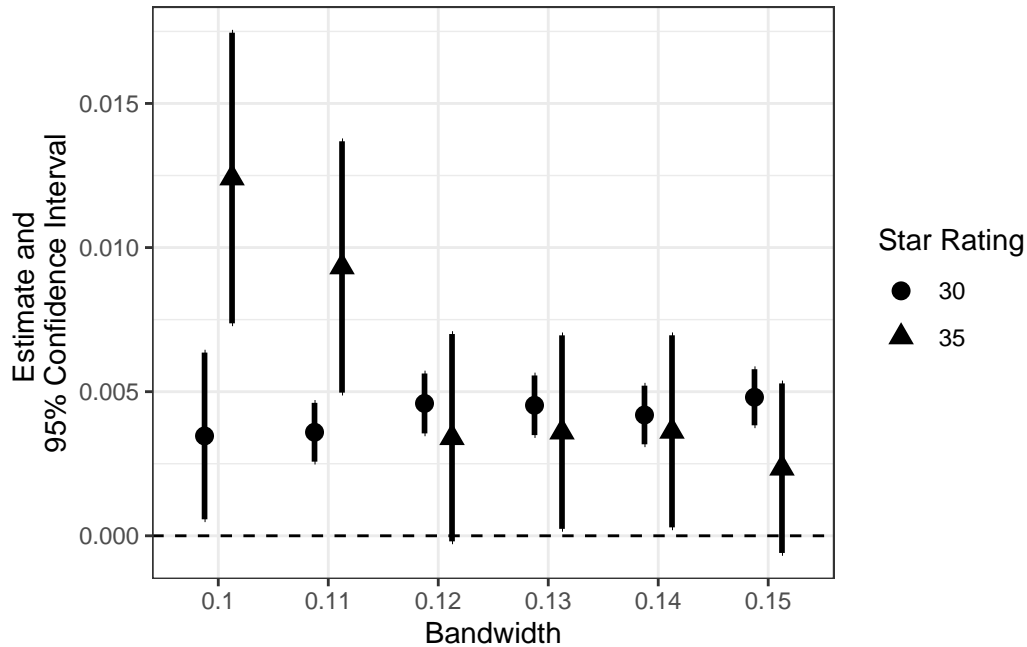

Figure 5: RD Estimates by Star Ratings and Bandwidth

4. 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 plots of plans around the 2.75 threshold and the 3.25 thresholds are presented in Figure 6. From the distribution around the 2.75 threshold, a couple of points stand out. First, there are very few plans in the left-side of the threshold, meaning very few plans are rounded down to 2.5 stars. We also see that the bulk of plans in this area have raw scores of around 2.85. Looking at the 3.25 threshold, we see a spike in plans just below the threshold.

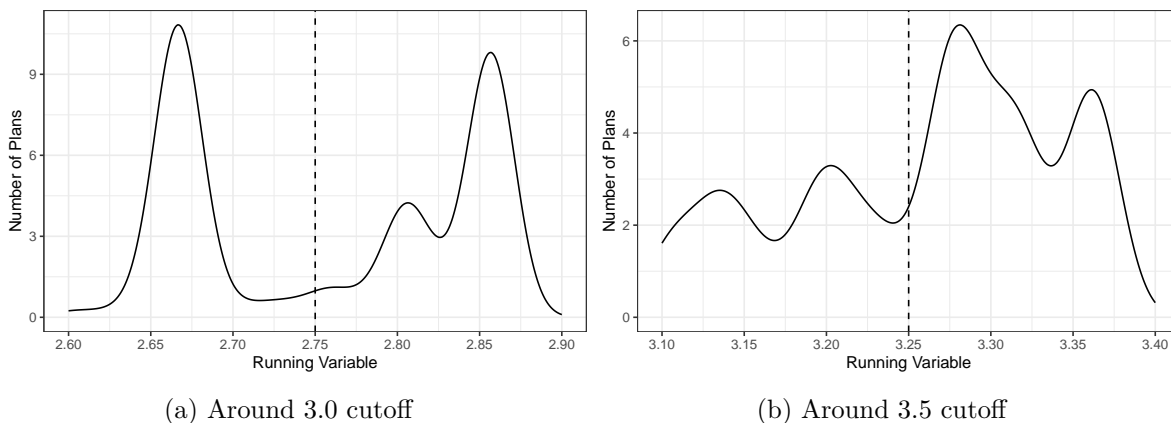
(a) Around 3.0 cutoff   (b) Around 3.5 cutoff

Figure 6: Density of Running Variable

Ultimately, while there are some apparent differences in the distribution of the raw scores above and below the threshold values, there doesn't appear to be much evidence of "manipulation" of the running variable. In other words, if insurers were manipulating this variable, then they would want to push their scores just above the threshold values in order to move into the higher rating category. We see no evidence of this in the data.

Note that there are much more formal ways of testing for these types of differences, but those are beyond what is expected for this class. If you are interested, please look into the `rdplotdensity` and `rddensity` commands, as well as the `DCdensity` command. I didn't explicitly ask for a formal test in the question, so I didn't do it as part of the answer key either; although there are examples of such formal tests in the slides.

5. 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.

For this, the standard approach is to use a "love plot" as we did earlier in the semester. In this case, I'm working with just two plan characteristics: 1) whether the plan is an "HMO" type, which is in the "plan_type" variable; and 2) whether the plan offers Part D (prescription drug) coverage. Balance plots are presented in Figure 7, where we see very large differences in plans rounded down to 2.5 versus those rounded up to 3.0 stars. This perhaps isn't surprising given the lack of data on plans rounded down to 2.5. Conversely, the covariates are more similar among those rounded up to 3.5 versus rounded down to 3 stars.
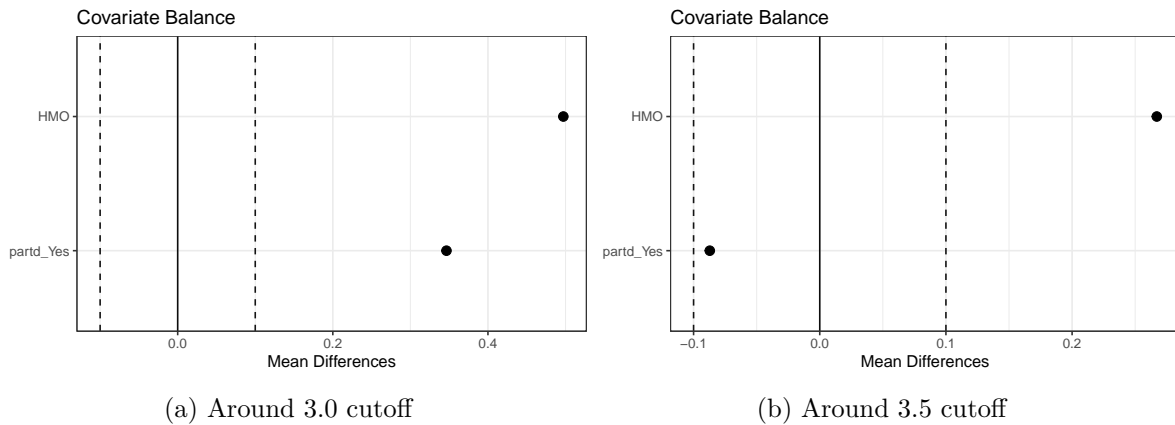

(a) Around 3.0 cutoff   (b) Around 3.5 cutoff

Figure 7: Balance Plots of Selected Plan Characteristcs

6. Summarize your findings from 1-5. What is the effect of increasing a star rating on enrollments? Briefly explain your results.

We estimate relatively small but statistically significant effects in going from 2.5 to 3.0 stars, with smaller and largely insignificant effects in moving from 3 to 3.5 stars. This is consistent with enrollees using ratings more to "avoid" low rated plans than to actively seek out higher rated plans.