Assignment3

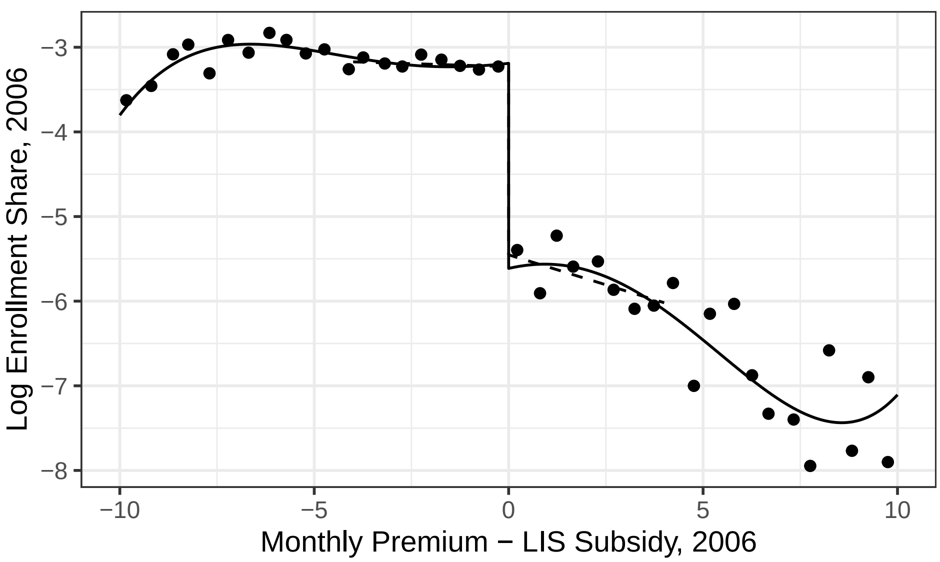
Jung Jae Kim

Questions

1. Recreate the table of descriptive statistics (Table 1) from Ericson (2014).

	2006	2007	2008	2009	2010
Mean monthly premium	$37	$40	$35	$30	$32
sd__premium	(12)	(17)	(19)	(5)	(9)
Mean deductible	$92	$114	$146	$253	$118
sd__deductible	(115)	(127)	(124)	(101)	(138)
Fraction enhanced benefit	0.43	0.43	0.58	0.03	0.69
...in the U.S.	0.00	0.76	0.98	1.00	0.97
...in the same state	0.00	0.76	0.98	1.00	0.97
Number of Unique Firms	51	38	16	5	6
Number of Plans	1429	658	202	68	107

2. Recreate Figure 3 from Ericson (2014).



3

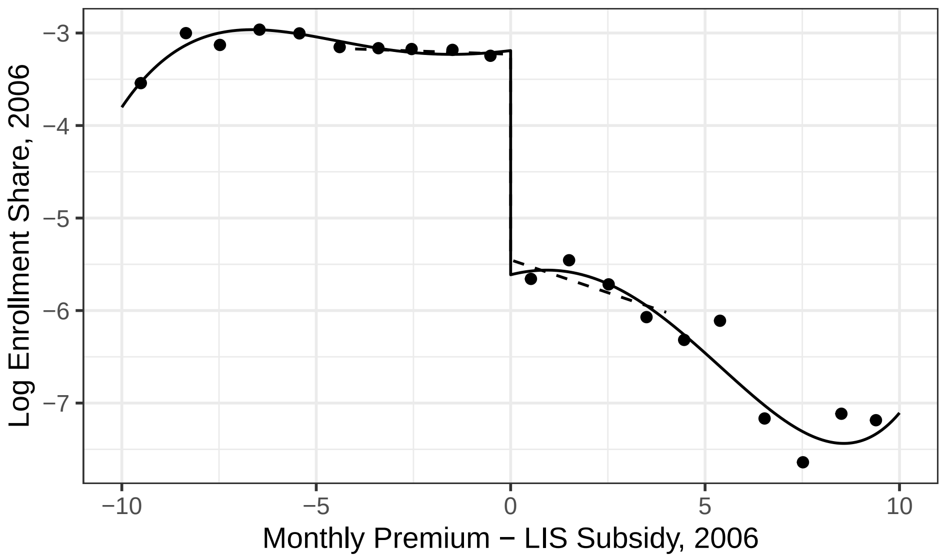
. Calonico, Cattaneo, and Titiunik (2015) discuss the appropriate partition size for binned scatterplots such as that in Figure 3 of Ericson (2014). More formally, denote by $\mathcal{P}_{-,n} = \{P_{-,j} : j = 1, 2, ...J_{-,n}\}$ and $\mathcal{P}_{+,n} = \{P_{+,j} : j = 1, 2, ...J_{+,n}\}$ the partitions of the support of the running variable x_i on the left and right (respectively) of the cutoff, $\bar{x}$. $P_{-,j}$ and $P_{+,n}$ denote the actual supports for each j partition of size $J_{-,n}$ and $J_{+,n}$, such that $[x_l, \bar{x}) = \bigcup_{j=1}^{J_{-,n}} P_{-,j}$ and $(\bar{x}, x_u] = \bigcup_{j=1}^{J_{+,n}} P_{+,j}$. Individual bins are denoted by $p_{-,j}$ and $p_{+,j}$. With this notation in hand, we can write the partitions $J_{-,n}$ and $J_{+,n}$ with equally-spaced bins as

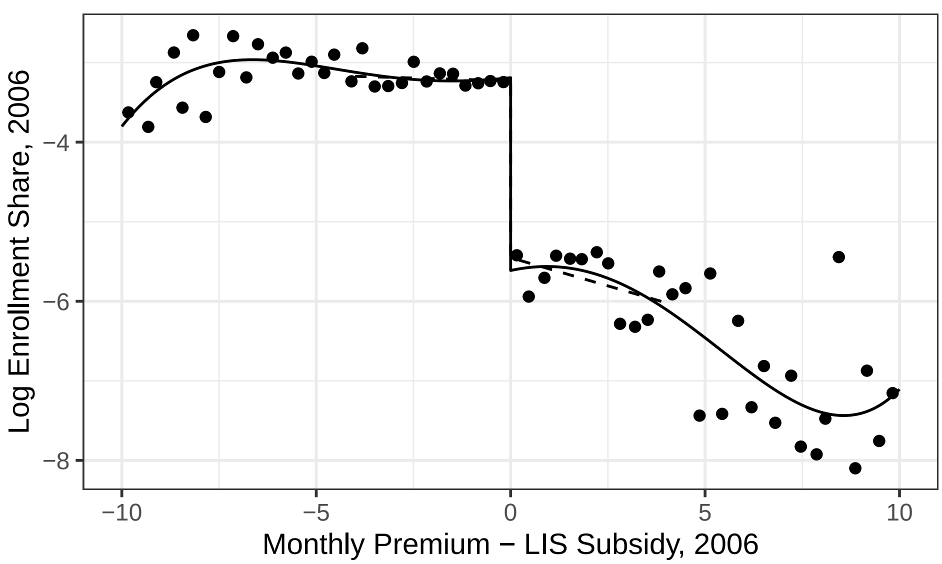
$$p_{-,j} = x_l + j \times \frac{\bar{x} - x_l}{J_{-,n}},$$

and

$$p_{+,j} = \bar{x} + j \times \frac{x_u - \bar{x}}{J_{+,n}}.$$

Recreate Figure 3 from Ericson (2014) using $J_{-,n} = J_{+,n} = 10$ and $J_{-,n} = J_{+,n} = 30$. Discuss your results and compare them to your figure in Part 2.





4. With the notation above, Calonico, Cattaneo, and Titiunik (2015) derive the optimal number of partitions for an evenly-spaced (ES) RD plot. They show that

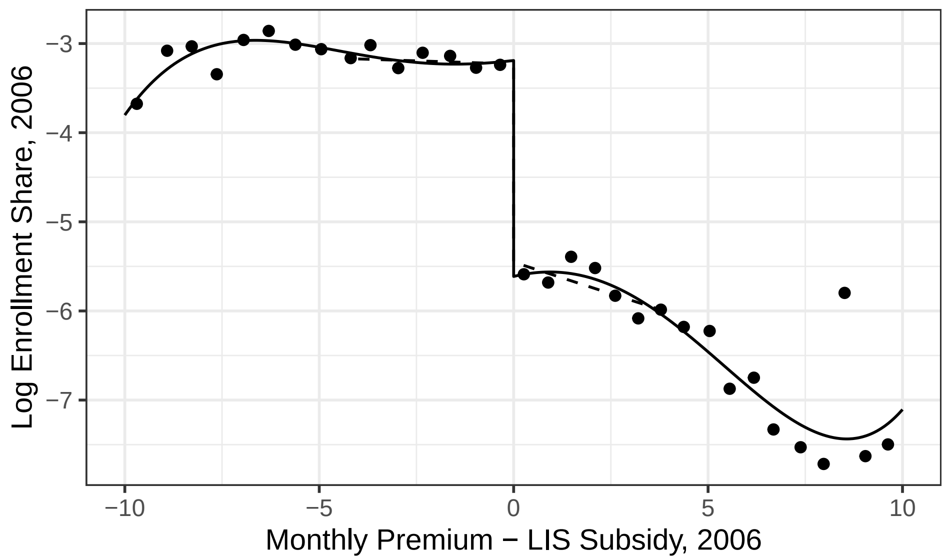
$$J_{ES,-,n} = \left\lceil \frac{V_-}{\mathcal{V}_{ES,-}} \frac{n}{\log(n)^2} \right\rceil$$

and

$$J_{ES,+,n} = \left\lceil \frac{V_+}{\mathcal{V}_{ES,+}} \frac{n}{\log(n)^2} \right\rceil,$$

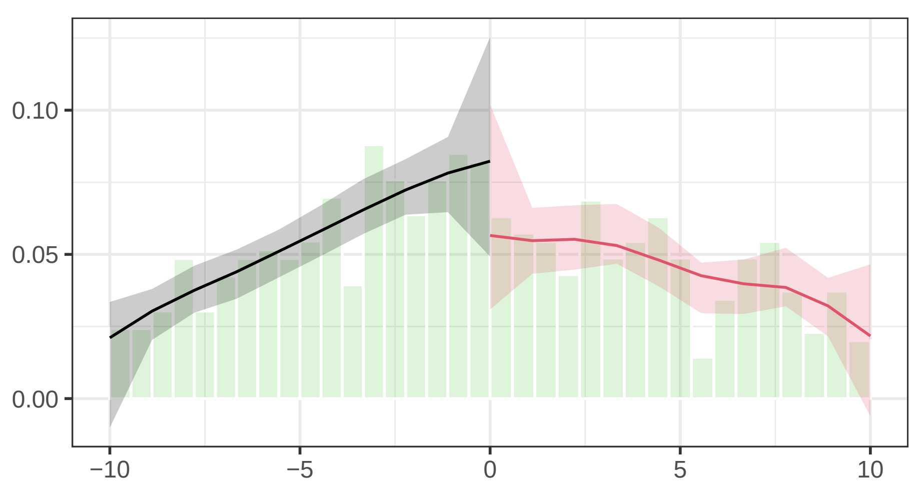
where V_- and V_+ denote the sample variance of the subsamples to the left and right of the cutoff and $\mathcal{V}_{ES,\cdot}$ is an integrated variance term derived in the paper. Use the `rdrobust` package in `R` (or `Stata` or `Python`) to find the optimal number of bins with an evenly-spaced binning strategy. Report this bin count and recreate your binned scatterplots from parts 2 and 3 based on the optimal bin number.

```
[1] "Optimal bin number is: J-=15 J+=17"
```



5. One key underlying assumption for RD design is that agents cannot precisely manipulate the running variable. While "precisely" is not very scientific, we can at least test for whether there appears to be a discrete jump in the running variable around the threshold. Evidence of such a jump may suggest that manipulation is present. Provide the results from the manipulation tests described in Cattaneo, Jansson, and Ma (2018). This test can be implemented with the `rddensity` package in `R`, `Stata`, or `Python`.

```
$Estl
Call: lpdensity

Sample size                                         324
Polynomial order for point estimation    (p=)     2
Order of derivative estimated            (v=)     1
Polynomial order for confidence interval (q=)     3
Kernel function                                   triangular
Scaling factor                                    0.545608108108108
Bandwidth method                                  user provided

Use summary(...) to show estimates.

$Estr
Call: lpdensity

Sample size                                         269
Polynomial order for point estimation    (p=)     2
Order of derivative estimated            (v=)     1
Polynomial order for confidence interval (q=)     3
Kernel function                                   triangular
Scaling factor                                    0.452702702702703
Bandwidth method                                  user provided

Use summary(...) to show estimates.

$Estplot
```



6. Recreate Panels A and B of Table 3 in Ericson (2014) using the same bandwidth of $4.00 but without any covariates.

	2006	2007	2008	2009	2010
Below Benchmark, 2006	2.224***	1.332***	0.902***	0.803**	0.677
	(0.283)	(0.267)	(0.248)	(0.362)	(0.481)
. . . Below Benchmark	-0.014	-0.077	-0.073	-0.170	-0.215**
	(0.032)	(0.088)	(0.116)	(0.105)	(0.088)
. . . Above Benchmark	-0.142*	-0.033	0.049	0.074	0.049
	(0.078)	(0.110)	(0.163)	(0.170)	(0.202)
Num.Obs.	306	299	298	246	212
R2	0.576	0.325	0.131	0.141	0.124

Note: ^^ * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

	2006	2007	2008	2009	2010
Below Benchmark, 2006	2.349***	1.206***	0.697*	0.238	0.152
	(0.279)	(0.387)	(0.394)	(0.516)	(0.633)
Num.Obs.	306	299	298	246	212
R2	0.577	0.327	0.137	0.163	0.140

Note: ^^ * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

7. Calonico, Cattaneo, and Farrell (2020) show that pre-existing optimal bandwidth calculations (such as those used in Ericson (2014)) are invalid for appropriate inference. They propose an alternative method to derive minimal coverage error (CE)-optimal bandwidths. Re-estimate your RD results using the CE-optimal bandwidth (`rdrobust` will do this for you) and compare the bandwidth and RD estimates to that in Table 3 of Ericson(2014).

	2006	2007	2008	2009	2010
coef	2.286	1.595	0.784	0.083	0.390
se	(0.502)	(0.634)	(0.568)	(0.65)	(0.831)
bw	0.748	0.675	0.964	0.770	0.771

	2006	2007	2008	2009	2010
coef	2.576	1.849	1.057	0.015	0.652
se	(0.593)	(0.755)	(0.742)	(0.771)	(1.175)
bw	1.022	0.979	1.188	0.994	0.922

8. Now let's extend the analysis in Section V of Ericson(2014) using IV. Use the presence of Part D low-income subsidy as an IV for market share to examine the effect of market share in 2006 on future premium changes.

	ivmodel
Dependent Var.:	log_premium
log_share	-9.729 (27.82)
Fixed-Effects:	----------
state	Yes
year	Yes
______________	______________
S.E.: Clustered	by: state
Observations	1,361
R2	-3,147.6
Within R2	-4,546.1

9. Discuss your findings and compare results from different binwidths and bandwidths. Compare your results in part 8 to the invest-then-harvest estimates from Table 4 in Ericson(2014).

Answers: Different binwidths and bandwidths give similar results.

10. Reflect on this assignment. What did you find most challenging? What did you find most surprising?

Answers: Replicating and doing robustness checks in RD design was helpful to better understand and implement RD using R.