Section 3: Medicare Advantage Quality and Regression Discontinuity

Ian McCarthy | Emory University
Econ 470 & HLTH 470

Table of contents

Background on Medicare Advantage and Quality

What is Medicare Advantage

- Private provision of health insurance for Medicare beneficiaries
- Medicare "replacement" plans
- It's just private insurance for Medicare folks

Medicare Advantage History

- Existed since 1980s, formalized in the 1990s, expanded in 2000s
- Medicare+Choice as part of Balanced Budget Act in 1997
- Largest expansion: Medicare Modernization Act in 2003 (also brought Medicare Part D)

Medicare Advantage Details

In its current form...

- Insurers submit plan details and a price needed to cover traditional Medicare ("bid")
- If approved, Medicare pays risk-adjusted bid *or* benchmark
- Bid $<$ benchmark, insurer gets a rebate
- Bid $>$ benchmark, insurer charges premium
- Seperate bidding for Part D

Medicare Advantage in Real Life

Let's take a look at the Medicare Advantage plan options...

Medicare Plan Finder

Medicare Advantage Quality Ratings

- Initial MA Star Ratings (2007)
- Overall rating introduced in 2009
- Complicated formula...

key point: ratings from several domains are averaged and then rounded

Regression Discontinuity in Theory

Intuition

Key intuition from RD:

Observations are **identical** just above/below threshold

Intuition

Highly relevant in "rule-based" world…

- School eligibility based on age cutoffs
- Program participation based on discrete income thresholds
- Performance scores rounded to nearest integer

Types of RD

1. Sharp regression discontinuity
    - those above the threshold guaranteed to participate

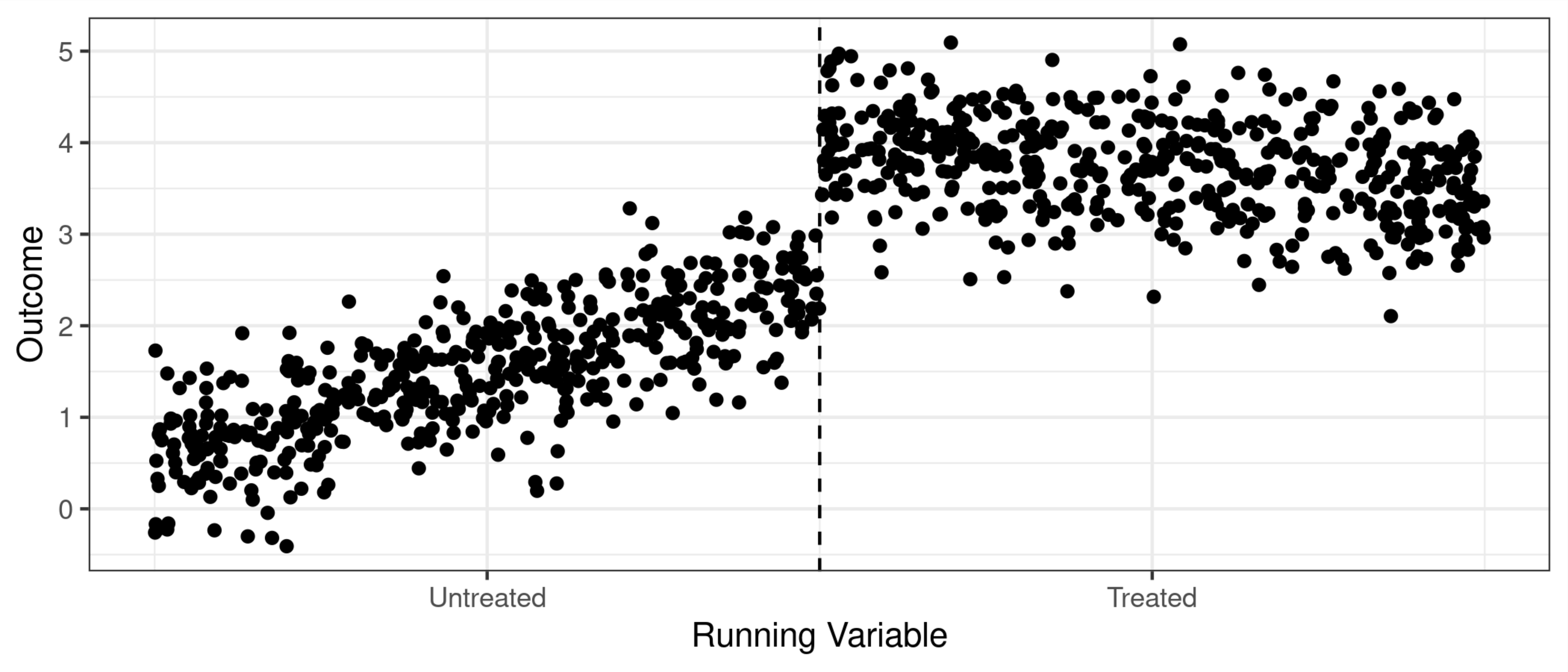
2. Fuzzy regression discontinuity
    - those above the threshold are eligible but may not participate

Sharp RD

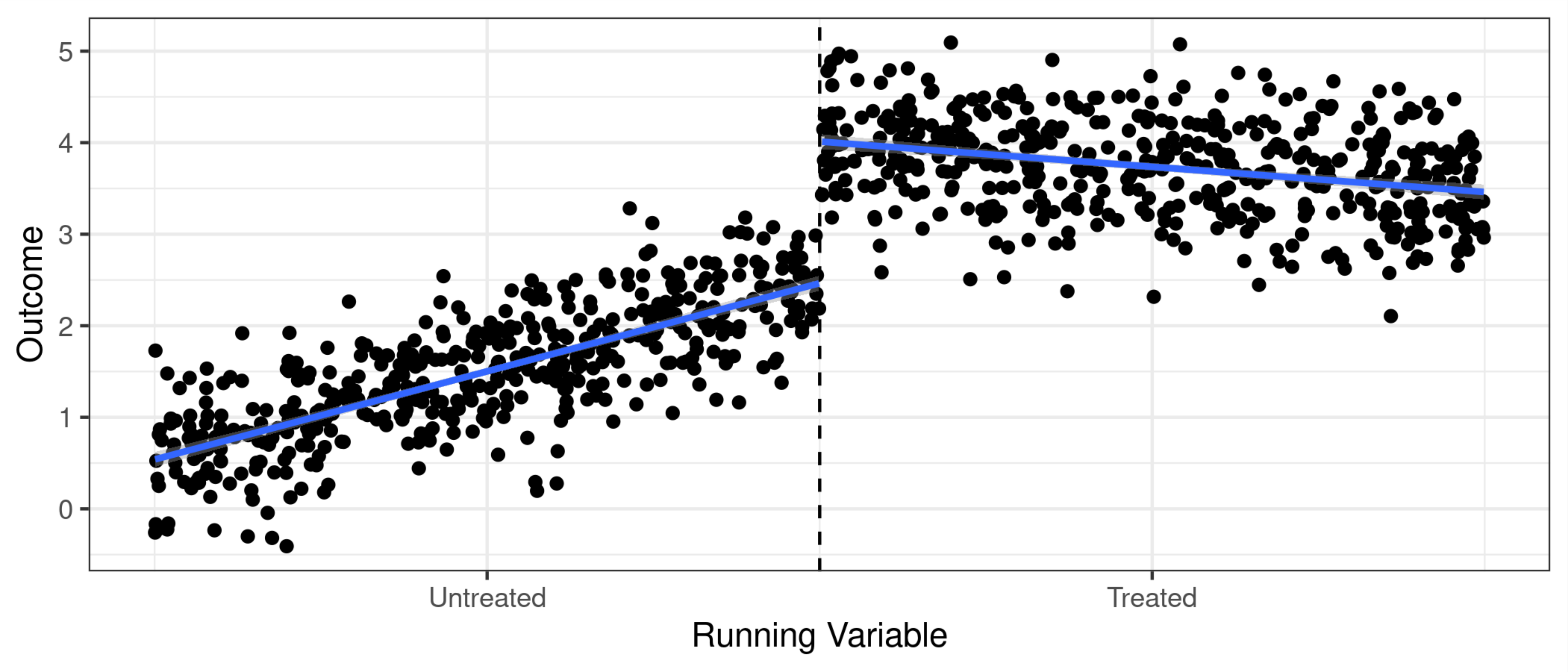
$$W_i = 1(x_i > c) = \begin{cases} 1 & \text{if} \quad x_i > c \\ 0 & \text{if} \quad x_i < c \end{cases}$$

- x is "forcing variable"
- c is the threshold value or cutoff point

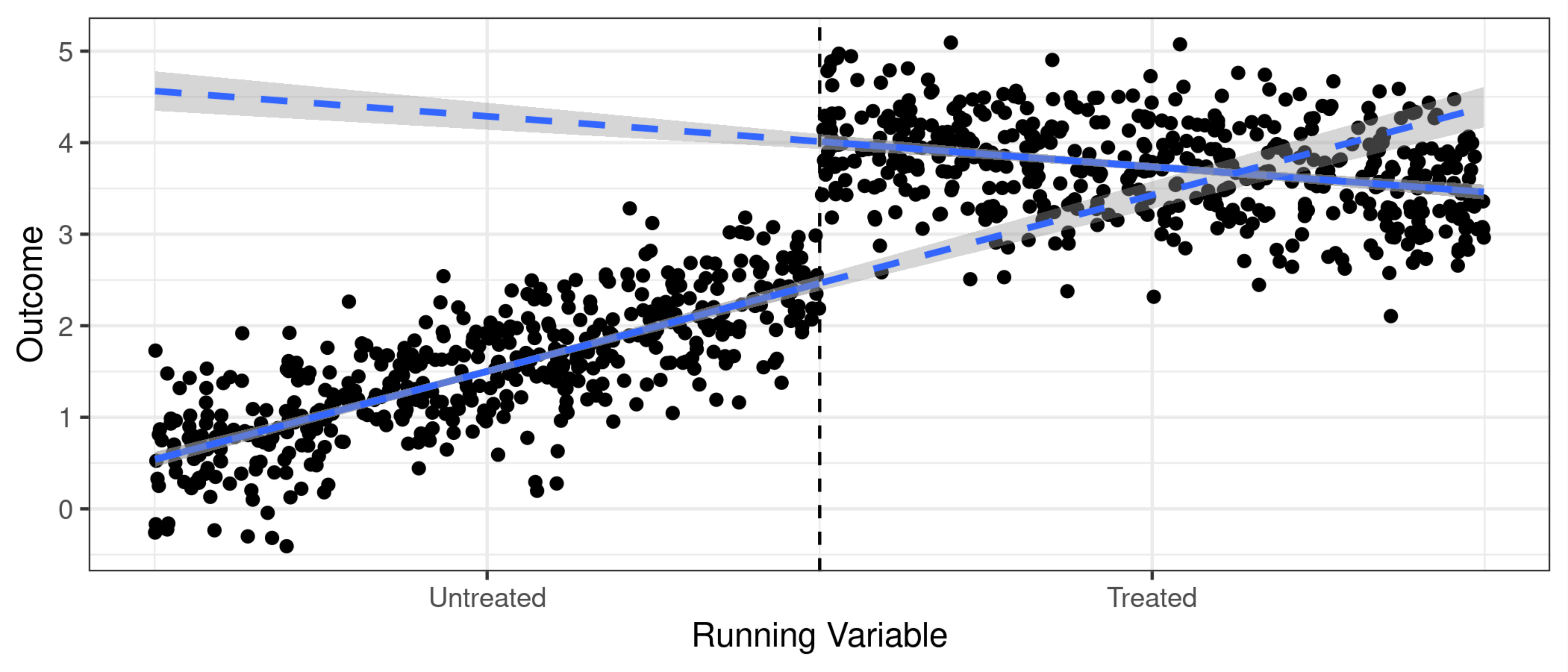
Sharp RD Scatterplot



Sharp RD Linear Predictions



Sharp RD Linear Predictions



Different averages

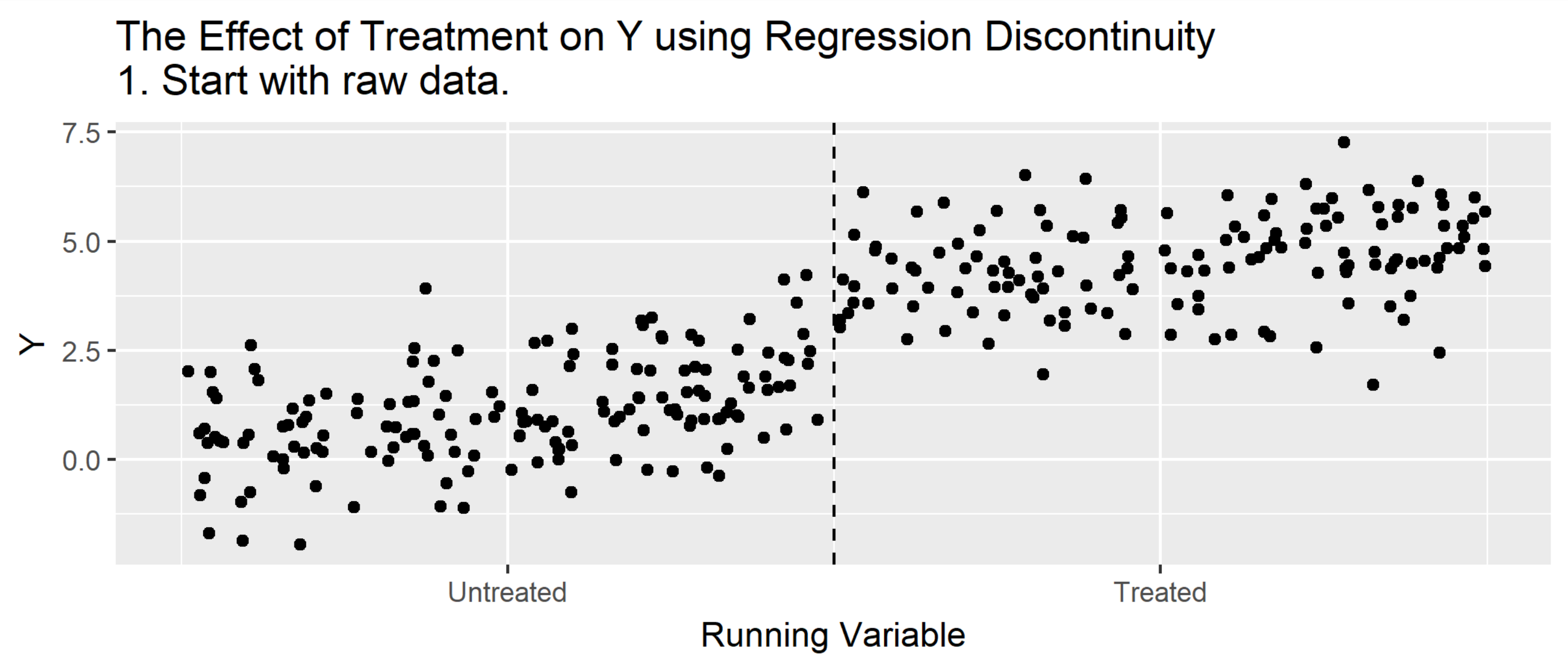
- Mean difference around threshold of 0.2, 3.97 - 2.25 = 1.72
- Mean overall difference, 3.74 - 1.49 = 2.25

More generally

- Running variable may affect outcome directly
- Focusing on area around cutoff does two things:

  1. Controls for running variable
  2. "Controls" for unobserved things correlated with running variable and outcome

Animations!


The Effect of Treatment on Y using Regression Discontinuity
1. Start with raw data.
7.5
5.0
2.5
0.0
Y
Untreated
Treated
Running Variable

Estimation

Goal is to estimate $E[Y(1)|X=c] - E[Y(0)|X=c]$

1. Trim to reasonable window around threshold ("bandwidth"), $X \in [c-h, c+h]$
2. Transform running variable, $\tilde{X} = X - c$
3. Estimate regressions...
    - Linear, same slope: $y = \alpha + \delta W + \beta \tilde{X} + \varepsilon$
    - Linear, different slope: $y = \alpha + \delta W + \beta \tilde{X} + \gamma W \tilde{X} + \varepsilon$
    - Nonlinear: add polynomials in $\tilde{X}$ and interactions $W\tilde{X}$

Regression Discontinuity in Practice

RDs "in the wild"

Most RD estimates follow a similar set of steps:

1. Investigate the running variable and show a jump at the discontinuity
2. Show clear graphical evidence of a change around the discontinuity
3. Overlay regression specification (and consider "Continuity-Based RD")
4. Explore sensitivity to bandwidths and orders of the polynomial
5. Conduct similar analyses with baseline covariates as outcomes
6. Explore sensitivity of results to inclusion of baseline covariates

Initial graphical evidence

Before presenting RD estimates, **any** good RD approach first highlights the discontinuity with a simple graph. We can do so by plotting the average outcomes within bins of the forcing variable (i.e., binned averages),

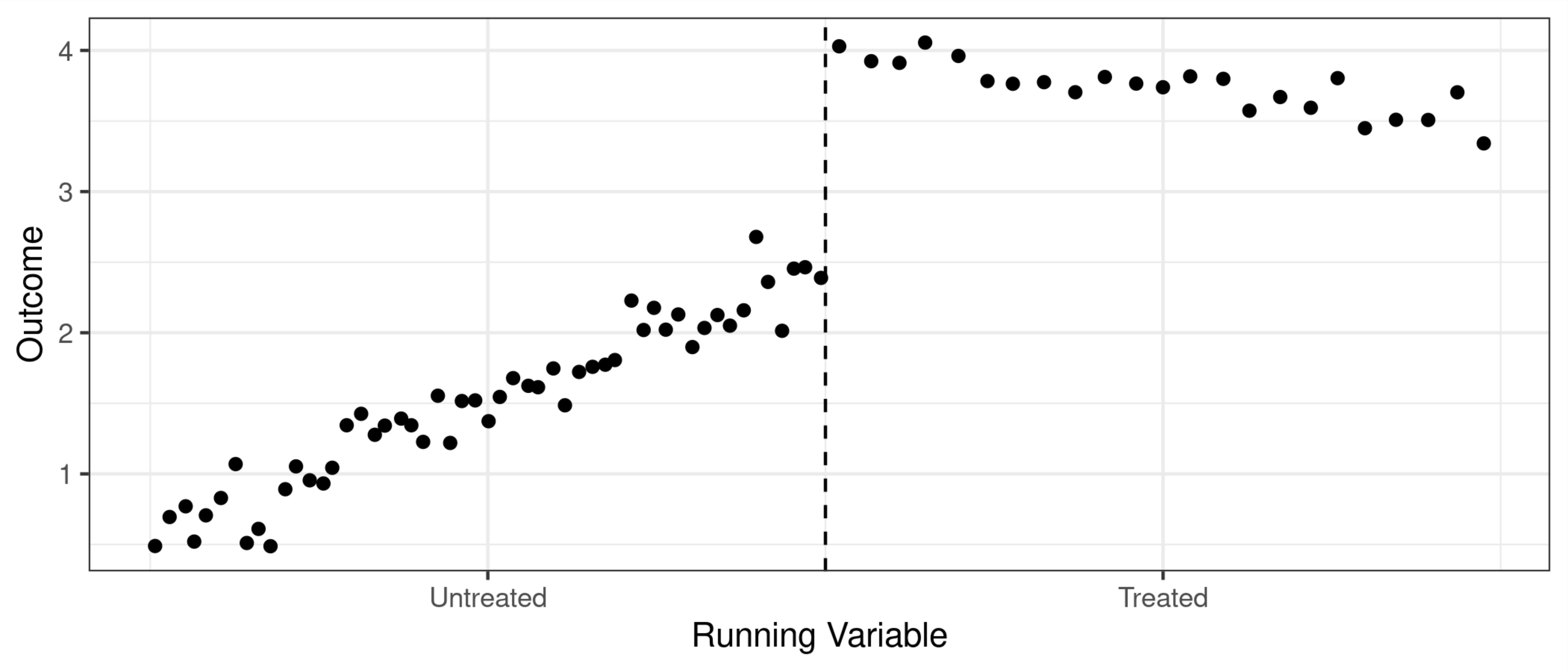
$$\bar{Y}_k = \frac{1}{N_k}\sum_{i=1}^{N} Y_i \times 1(b_k < X_i \leq b_{k+1}).$$

The binned averages helps to remove noise in the graph and can provide a cleaner look at the data. Just make sure that no bin includes observations above and below the cutoff!

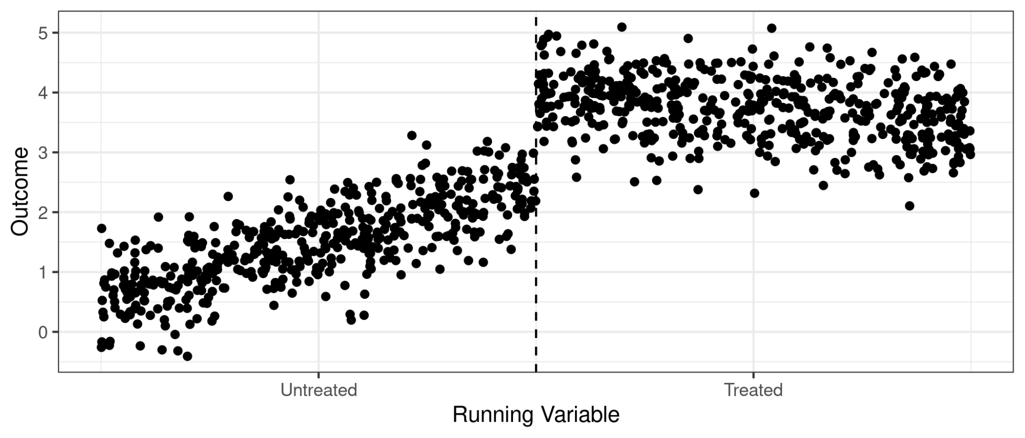
Binned average calculation

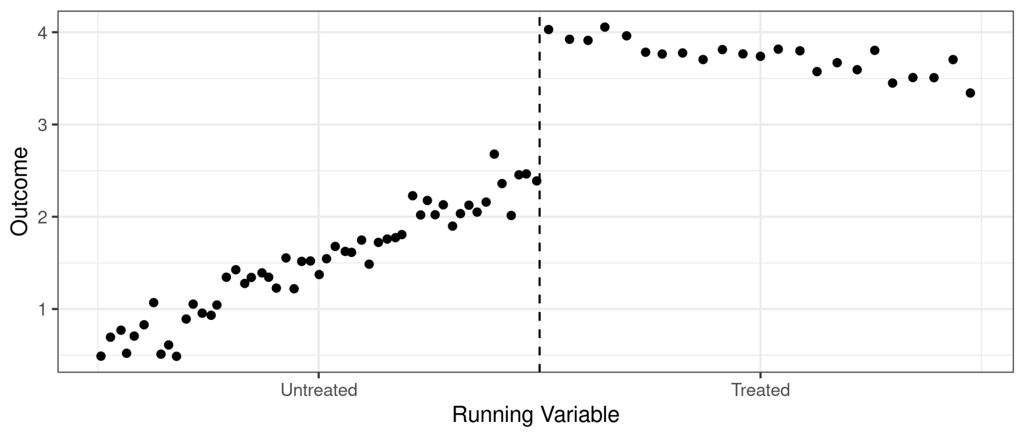
```r
library(rdrobust)
rd.result <- rdplot(rd.dat$Y, rd.dat$X,
                    c=1,
                    title="RD Plot with Binned Average",
                    x.label="Running Variable",
                    y.label="Outcome")
```

Binned average plot



With and without binning





Kernels?

Some RD estimates talk about "kernel weighting" to assign more weight to observations closer to the threshold and less weight to observations further from the threshold.

Kernels

$$\hat{\mu}_+(x) = \frac{\sum_{i:X_i<c} Y_i \times K\left(\frac{X_i - x}{h}\right)}{\sum_{i:X_i<c} K\left(\frac{X_i - x}{h}\right)},$$

and

$$\hat{\mu}_-(x) = \frac{\sum_{i:X_i\geq c} Y_i \times K\left(\frac{X_i - x}{h}\right)}{\sum_{i:X_i\geq c} K\left(\frac{X_i - x}{h}\right)},$$

where $K(u)$ is a kernel that assigns weight to observations based on the distance from u. A rectagular kernel is such that $K(u) = 1/2$ for $u \in (-1, 1)$ and 0 elsewhere.

Kernels and regression

- Local linear regression (regression within the pre-specified bandwidth) is a kernel weighted regression with a uniform (or rectangular) kernel.
- Could use more complicated kernels for a fully nonparametric approach, but these don't work well around the RD cutoff values.
- Polynomial

Some practical concerns

- Bin size for plots
- Selecting bandwidth, h
- Check for sorting around threshold (e.g., gaming)
- Covariate balance (love plots around threshold)
- Should we control for other covariates?
- Sensitivity to polynomial specification

Selecting "bin" width

1. Dummy variables: Create dummies for each bin, regress the outcome on the set of all dummies R_r^2, repeat with double the number of bins and find r-square value R_u^2, form F-stat, $\frac{R_u^2 - R_r^2}{1 - R_u^2} \times \frac{n - K - 1}{K}$.

2. Interaction terms: Include interactions between dummies and the running variable, joint F-test for the interaction terms

If F-test suggests significance, then we have too few bins and need to narrow the bin width.

Selecting bandwidth in local linear regression

The bandwidth is a "tuning parameter"

- High h means high bias but lower variance (use more of the data, closer to OLS)
- Low hh means low bias but higher variance (use less data, more focused around discontinuity)

Represent bias-variance tradeoff with the mean-square error,

$$MSE(h) = E[(\hat{\tau}_h - \tau_{RD})^2] = (E[\hat{\tau}_h - \tau_{RD}])^2 + V(\hat{\tau}_h).$$

Selecting bandwidth

In the RD case, we have two different mean-square error terms:

1. "From above", $MSE_+(h) = E[(\hat{\mu}_+(c,h) - E[Y_i(1)|X_i = c])^2]$
2. "From below", $MSE_-(h) = E[(\hat{\mu}_-(c,h) - E[Y_i(0)|X_i = c])^2]$

Goal is to find h that minimizes these values, but we don't know the true $E[Y(1)|X = c]$ and $E[Y(0)|X = c]$. So we have two approaches:

1. Use **cross-validation** to choose h
2. Explicitly solve for optimal bandwidth

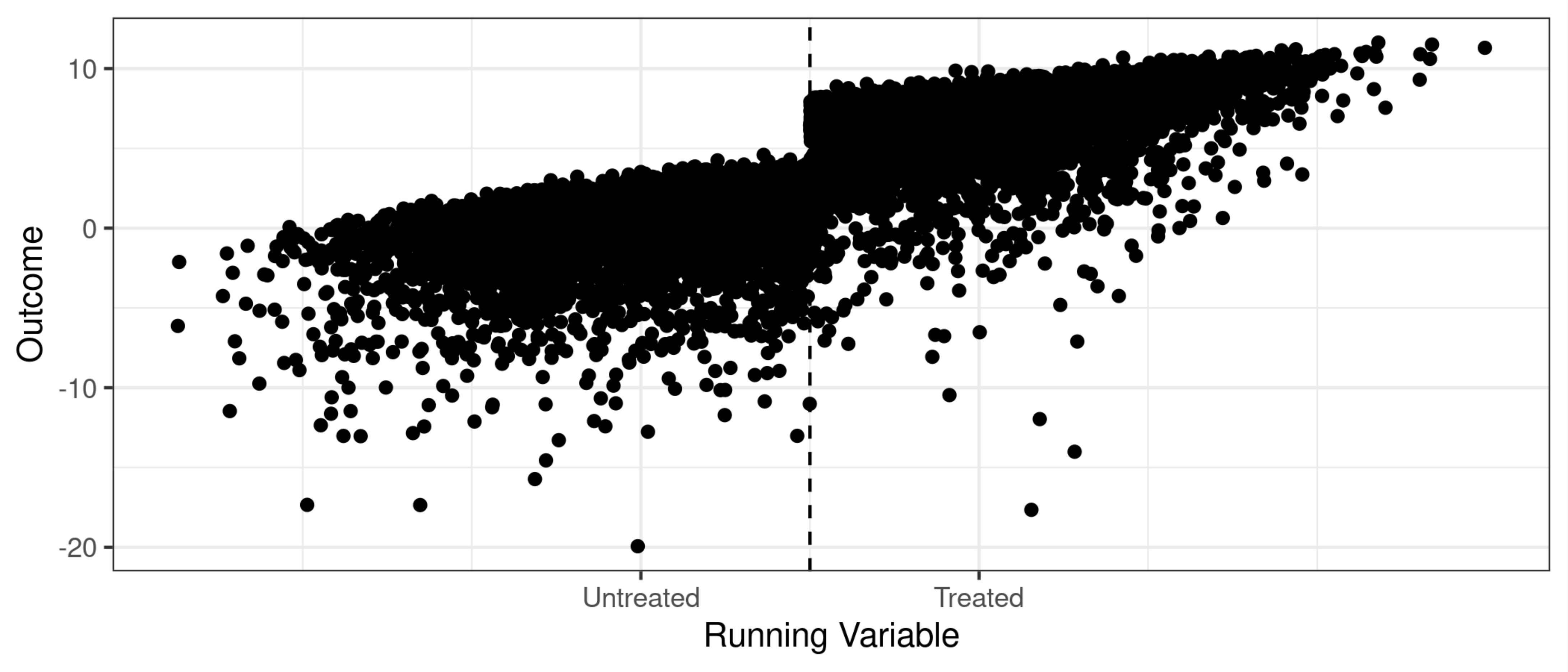
Cross-validation

Essentially a series of "leave-one-out" estimates:

1. Pick an h
2. Run regression, leaving out observation i. If i is to the left of the threshold, we estimate regression for observations within $X_i - h$, and conversely $X_i + h$ if i is to the right of the threshold.
3. Predicted $\hat{Y}_i$ at X_i (out of sample prediction for the left out observation)
4. Do this for all i, and form $CV(h) = \frac{1}{N} \sum (Y_i - \hat{Y}_i)^2$

Select h with lowest $CV(h)$ value.

Back to simulated data



Back to simulated data

```r
ols <- lm(Y~X+W, data=rd.dat2)

rd.dat3 <- rd.dat2 %>%
  mutate(x_dev = X-1) %>%
  filter( (X>0.8 & X <1.2) )
rd <- lm(Y~x_dev + W, data=rd.dat3)
```

- True effect: 4
- Standard linear regression with same slopes: 3.44
- RD (linear with same slopes): 3.84

Fuzzy Regression Discontinuity

Fuzzy RD

"Fuzzy" just means that assignment isn't guaranteed based on the running variable. For example, maybe students are much more likely to get a scholarship past some threshold SAT score, but it remains possible for students below the threshold to still get the scholarship.

- Discontinuity reflects a jump in the probability of treatment
- Other RD assumptions still required (namely, can't manipulate running variable around the threshold)

Fuzzy RD is IV

In practice, fuzzy RD is employed as an instrumental variables estimator

- Difference in outcomes among those above and below the discontinuity divided by the difference in treatment probabilities for those above and below the discontinuity,
  $E[Y_i|W_i = 1] - E[Y_i|W_i = 0] = \frac{E[Y_i|x_i \geq c] - E[Y_i|x_i < c]}{E[W_i|x_i \geq c] - E[W_i|x_i < c]}$
- Indicator for $x_i \geq c$ is an instrument for treatment status, W_i.

Medicare Advantage Data

Medicare Advantage

- Recall the Medicare Advantage repository, Medicare Advantage GitHub repository
- Now we need to work with the full dataset

Full MA Data

```
ma.data <- read_rds(here("data/final_ma_data.rds"))
```

Summary stats

Focus on enrollments and star ratings:

```
stargazer(as.data.frame(ma.data %>% select(avg_enrollment, avg_eligibles, Star_Rating)), type="html")
```

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
planid	895,495	35.849	69.789	1	4	41	999
avg_enrollment	204,173	397.404	1,578.641	11.000	31.250	231.667	63,234.080
avg_eligibles	749,267	42,587.920	98,741.840	11.571	3,798.000	37,116.570	1,355,734.000
Star_Rating	448,793	3.307	0.783	1.500	2.500	4.000	5.000

Clean the data

Limit to plans with:

- Observed enrollments, > 10
- First year of star rating (2009)
- Observed star rating

```
ma.data.clean <- ma.data %>%
  filter(!is.na(avg_enrollment) & year==2009 & !is.na(partc_score))
```

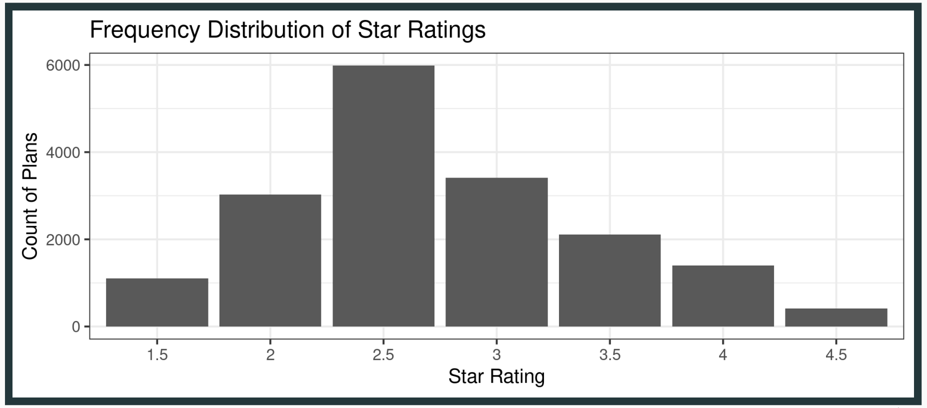
Calculate raw average rating

```
ma.data.clean <- ma.data.clean %>%
  mutate(raw_rating=rowMeans(
    cbind(breastcancer_screen,rectalcancer_screen,cv_cholscreen,diabetes_cholscreen,
          glaucoma_test,monitoring,flu_vaccine,pn_vaccine,physical_health,
          mental_health,osteo_test,physical_monitor,primaryaccess,
          hospital_followup,depression_followup,nodelays,carequickly,
          overallrating_care,overallrating_plan,calltime,
          doctor_communicate,customer_service,osteo_manage,
          diabetes_eye,diabetes_kidney,diabetes_bloodsugar,
          diabetes_chol,antidepressant,bloodpressure,ra_manage,
          copd_test,betablocker,bladder,falling,appeals_timely,
          appeals_review),
    na.rm=T)) %>%
  select(contractid, planid, fips, avg_enrollment, first_enrollment,
         last_enrollment, state, county, raw_rating, partc_score,
         avg_eligibles, avg_enrolled, premium_partc, risk_ab, Star_Rating,
         bid, avg_ffscost, ma_rate)
```

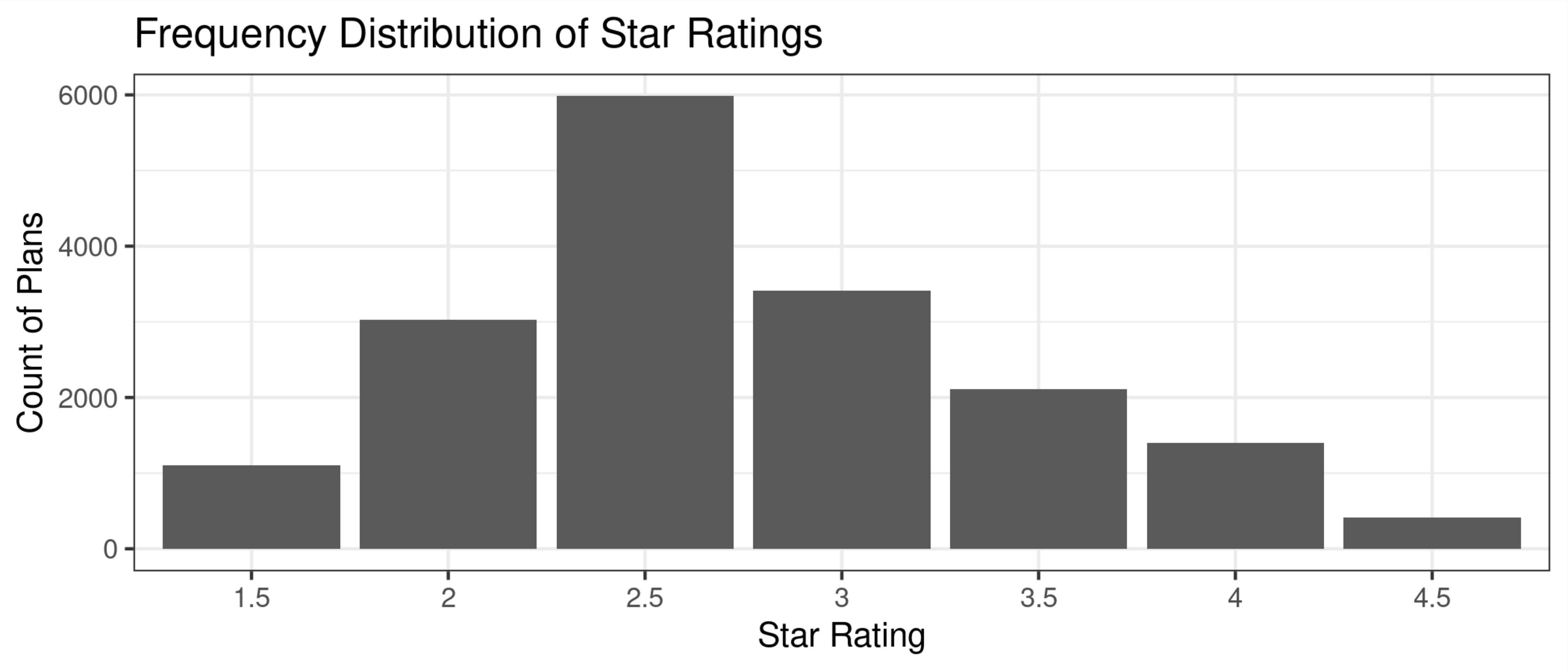
Quality and Insurance Choice

Distribution of star ratings

```
ma.data.clean %>%
  ggplot(aes(x=as.factor(Star_Rating))) +
  geom_bar() +
  labs(
    x="Star Rating",
    y="Count of Plans",
    title="Frequency Distribution of Star Ratings"
  ) + theme_bw()
```



Distribution of star ratings



Enrollments and star ratings

```
##
## Call:
## lm(formula = avg_enrollment ~ factor(Star_Rating), data = ma.data.clean)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##   -627   -388   -214    -51  41908
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)               87.31      43.32   2.016  0.04387 *
## factor(Star_Rating)2      32.75      50.62   0.647  0.51758
## factor(Star_Rating)2.5   194.65      47.15   4.128 3.67e-05 ***
## factor(Star_Rating)3     433.95      49.84   8.707  < 2e-16 ***
## factor(Star_Rating)3.5   470.91      53.47   8.808  < 2e-16 ***
## factor(Star_Rating)4     552.30      57.91   9.538  < 2e-16 ***
## factor(Star_Rating)4.5   272.36      82.68   3.294  0.00099 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1440 on 17451 degrees of freedom
## Multiple R-squared:  0.01559,    Adjusted R-squared:  0.01526
## F-statistic: 46.07 on 6 and 17451 DF,  p-value: < 2.2e-16
```

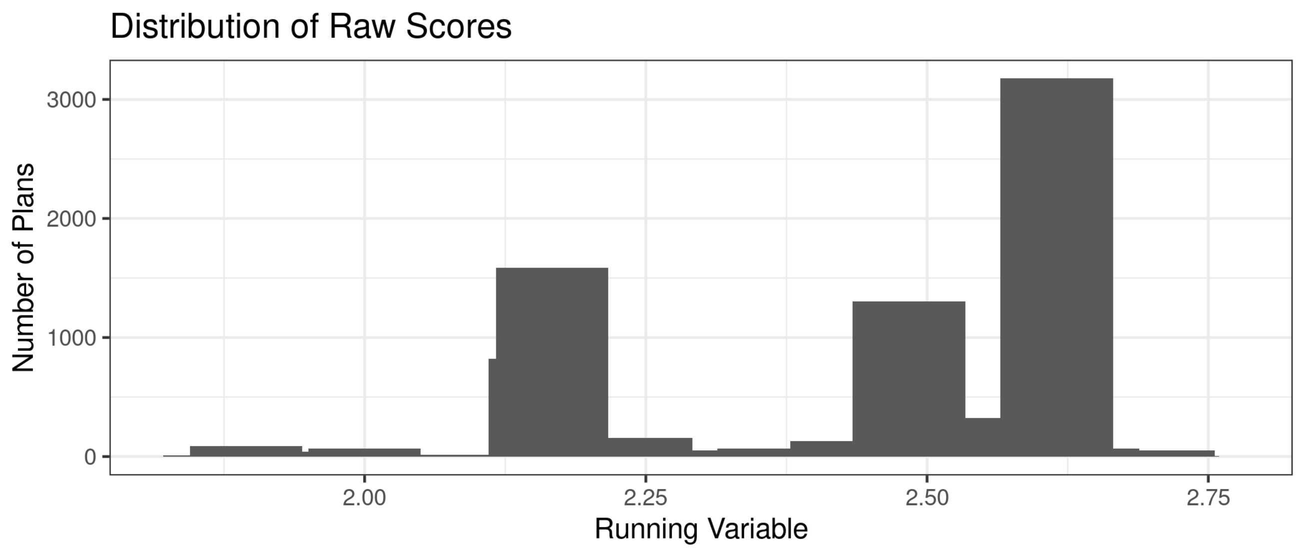
Problems

- Certainly not the effect of a higher rating...
- Lots of things unobserved, like
    - actual quality
    - perceived quality
    - prices

Effect of 3-star rating

```r
ma.rd1 <- ma.data.clean %>%
  filter(Star_Rating==2 | Star_Rating==2.5)
```

```r
ma.rd1 %>% ggplot(aes(x=raw_rating)) +
  geom_bar(width=.1) + theme_bw() +
  labs(
    x="Running Variable",
    y="Number of Plans",
    title="Distribution of Raw Scores"
  )
```



Note about scores

CMS does more than just an average...

- variance across individual metrics
- high variance is punished, low variance rewarded

RD estimates

```r
ma.rd1 <- ma.rd1 %>%
  mutate(score = raw_rating - 2.25,
         treat = (Star_Rating==2.5),
         window1 = (score>-.175 & score<.175),
         window2 = (score>-.125 & score<.125),
         mkt_share = avg_enrollment/avg_eligibles,
         ln_share = log(mkt_share))
star25.1 <- lm(mkt_share ~ treat, data=ma.rd1)
star25.2 <- lm(mkt_share ~ score + treat, data= (ma.rd1 %>% filter(window1==TRUE)))
star25.3 <- lm(mkt_share ~ score + treat, data= (ma.rd1 %>% filter(window2==TRUE)))
est1 <- as.numeric(star25.3$coef[3])
est2 <- as.numeric(star25.1$coef[2])
```

RD estimates

	Market Share		
	(1)	(2)	(3)
Raw Score		-0.044***	-0.090***
		(0.009)	(0.011)
Treatment	0.005***	0.009***	0.012***
	(0.0004)	(0.002)	(0.002)
Bandwith	0.5	0.175	0.125
Observations	9,006	3,095	2,937

Interpretation

These estimates suggest a 1.2% increase in market shares among 2.5-star plans versus an "identical" 2-star plan, compared to an OLS estimate of 0.5%.

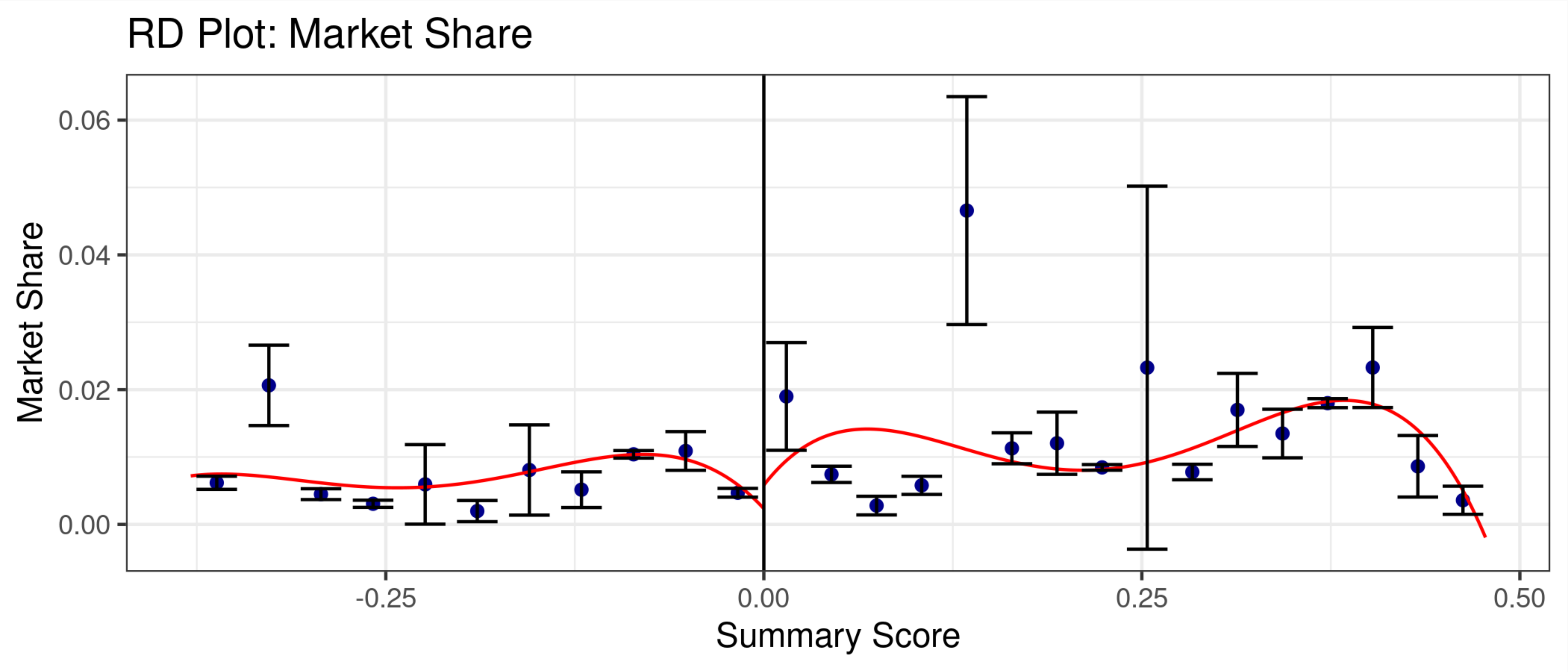
Built-in RD packages

```r
library(rdrobust)
rdplot(y=ma.rd1$mkt_share, x=ma.rd1$score, binselect="es", ci=95,
       title="RD Plot: Market Share", x.label="Summary Score",
       y.label="Market Share")
```

```
## [1] "Mass points detected in the running variable."
```

RD Plot

```
## [1] "Mass points detected in the running variable."
```



Estimates from RD package

```
est1 <- rdrobust(y=ma.rd1$mkt_share, x=ma.rd1$score)
```

```
## [1] "Mass points detected in the running variable."
```

```
summary(est1)
```

```
## Call: rdrobust
##
## Number of Obs.                  9006
## BW type                        mserd
## Kernel                    Triangular
## VCE method                        NN
##
## Number of Obs.                  3024          5982
## Eff. Number of Obs.              275           136
## Order est. (p)                     1             1
## Order bias  (q)                    2             2
## BW est. (h)                    0.065         0.065
## BW bias (b)                    0.180         0.180
## rho (h/b)                      0.362         0.362
## Unique Obs.                       20            46
##
```