Module 3: Medicare Advantage Quality and Regression Discontinuity

Part 2: Regression Discontinuity

Ian McCarthy | Emory University
Econ 470 & HLTH 470

Motivation

- Basic idea: Observations are **identical** just above/below threshold
- Some motivation from Causal Inference: The Mixtape

Motivation

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

Required elements

1. Score
2. Cutoff
3. Treatment

Types of RD

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

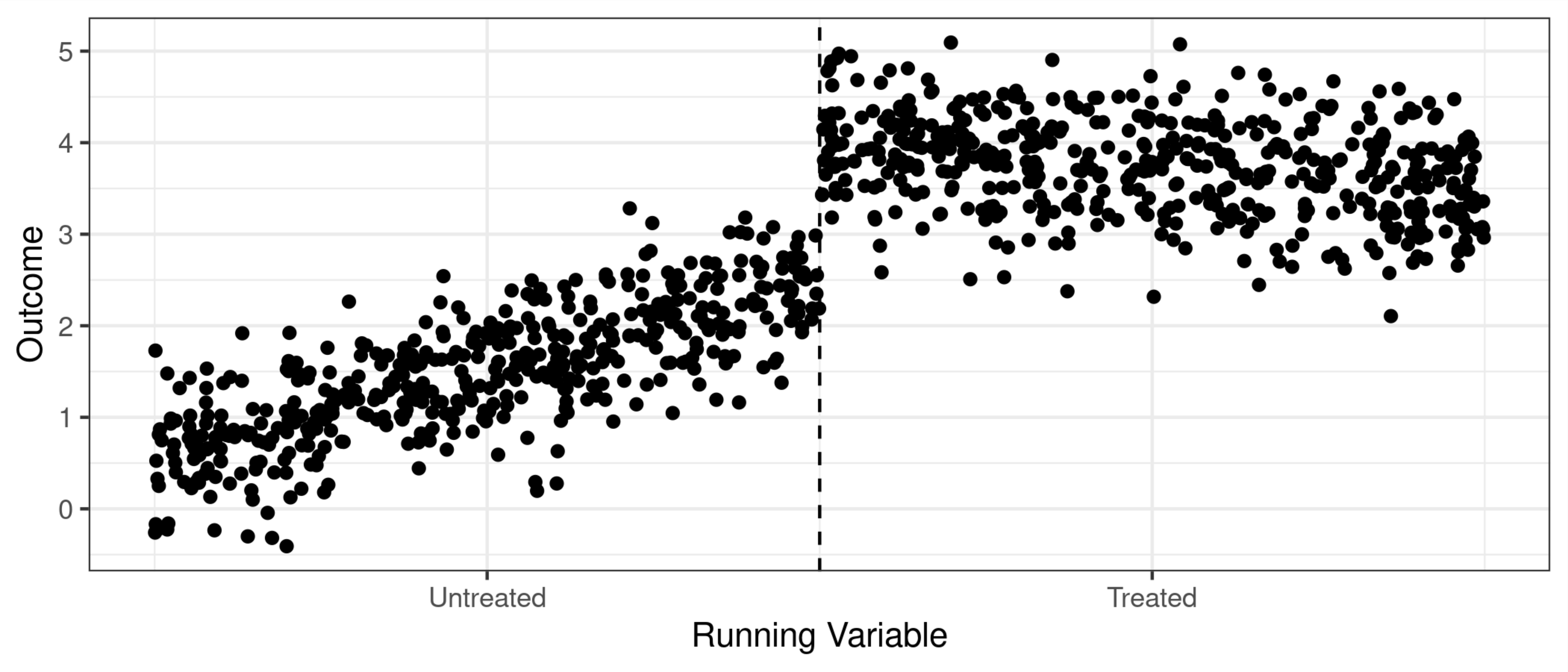
2. Fuzzy regression discontinuity
   - those above the threshold are eligible but may not participate

Sharp RD

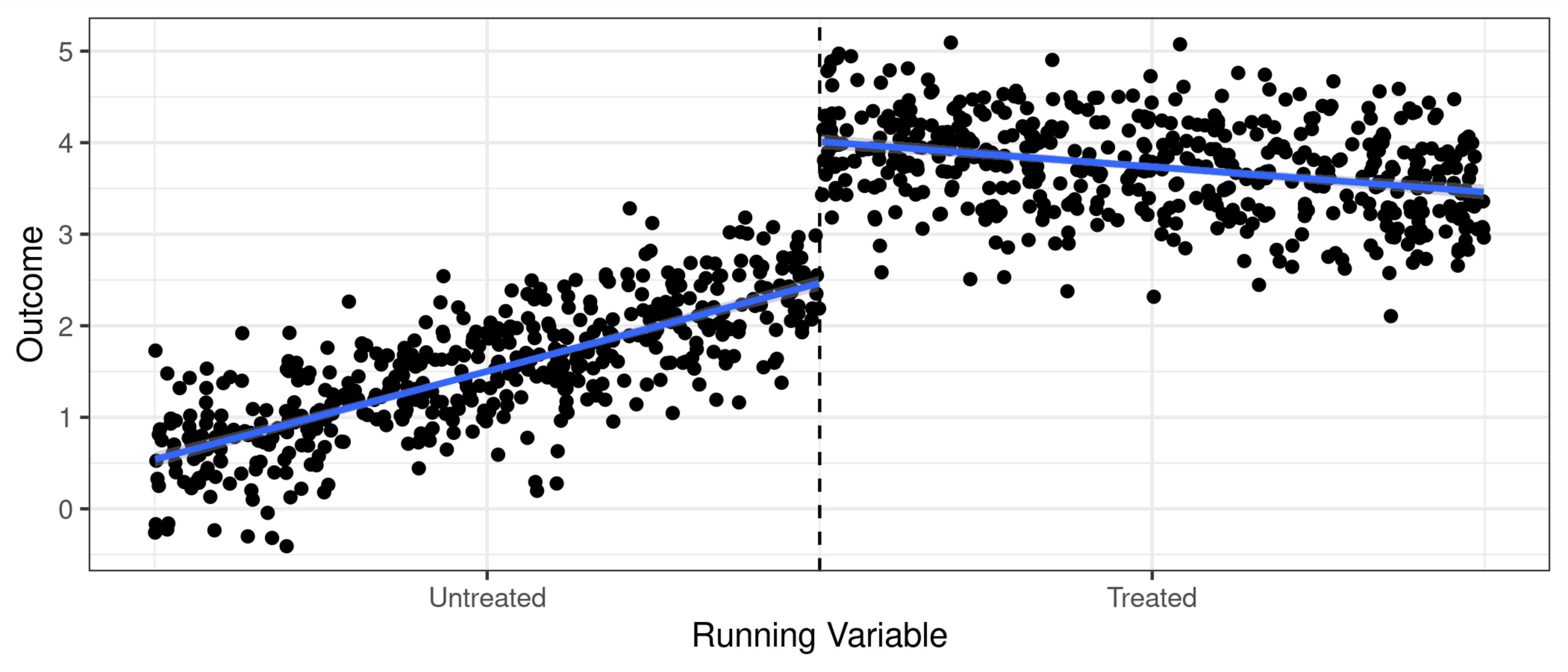
$$W_i = 1(x_i > c) = \begin{cases} 1 & \text{if} \quad x_i > c \\ 0 & \text{if} \quad x_i < c \end{cases}$$

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

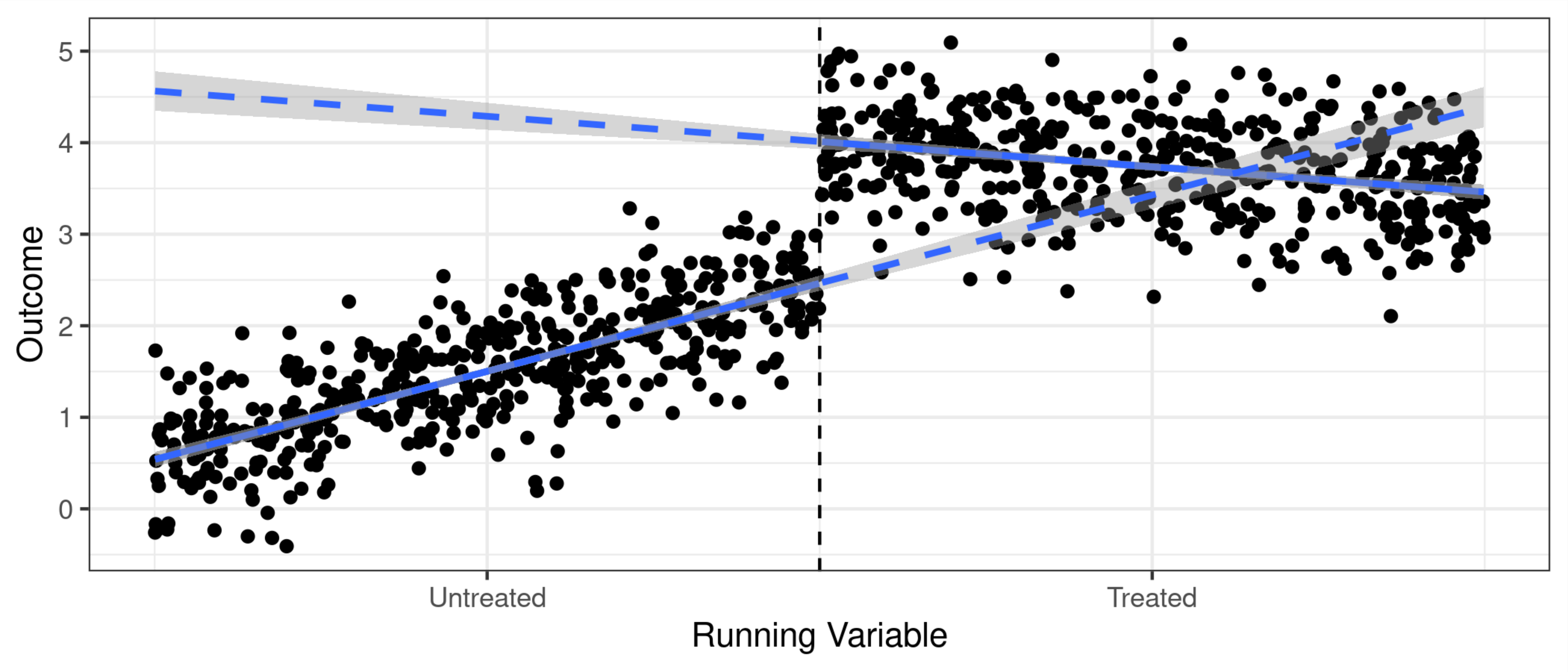
Sharp RD Scatterplot



Sharp RD Linear Predictions



Sharp RD Linear Predictions



Different averages

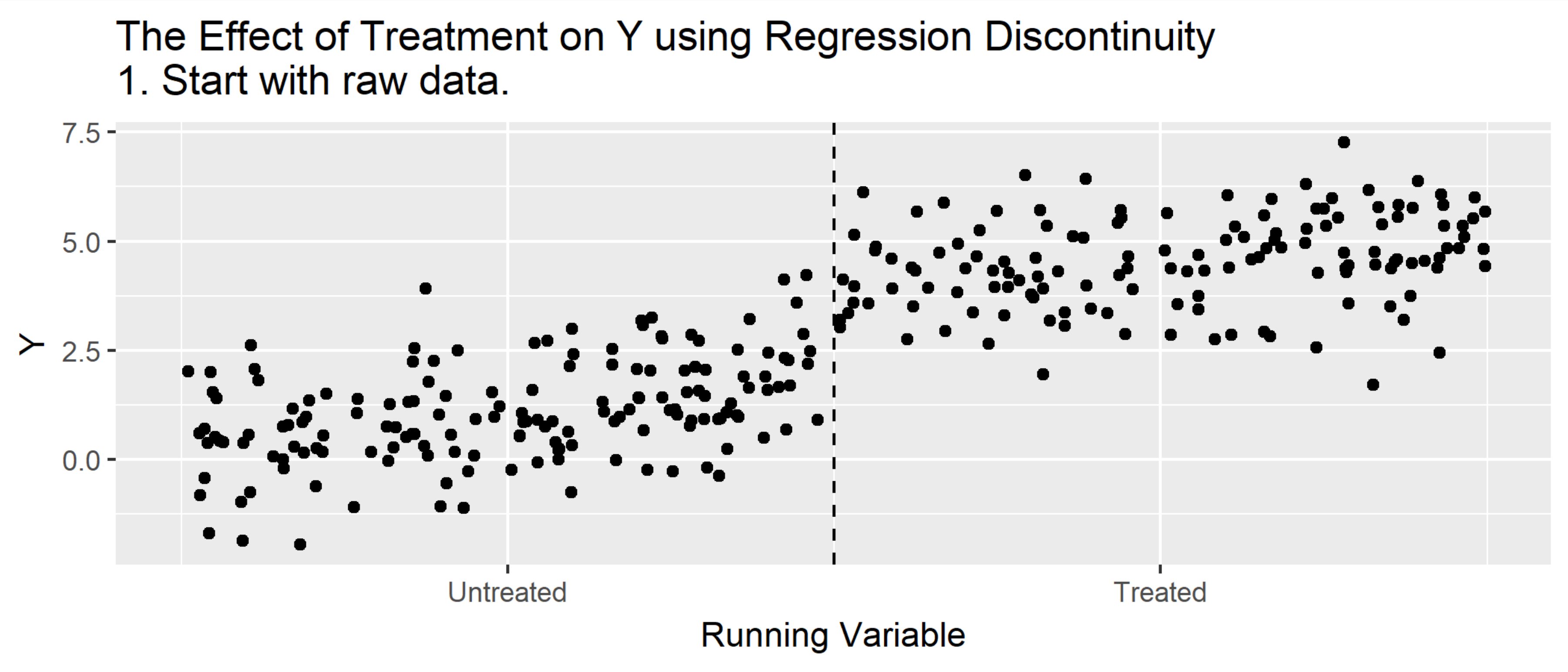
- Mean difference around threshold of 0.2, 3.97 - 2.25 = 1.72
- Mean overall difference, 3.74 - 1.49 = 2.25

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: $y = \alpha + \delta D + \beta\tilde{X} + \varepsilon$
   - Linear, different slope: $y = \alpha + \delta D + \beta\tilde{X} + \gamma W\tilde{X} + \varepsilon$
   - 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. Show clear graphical evidence of a change around the discontinuity (bin scatter)
2. Balance above/below threshold (use baseline covariates as outcomes)
3. Manipulation tests
4. RD estimates
5. Sensitivity and robustness:
    - Bandwidths
    - Order of polynomial
    - Inclusion of covariates

1. 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),

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

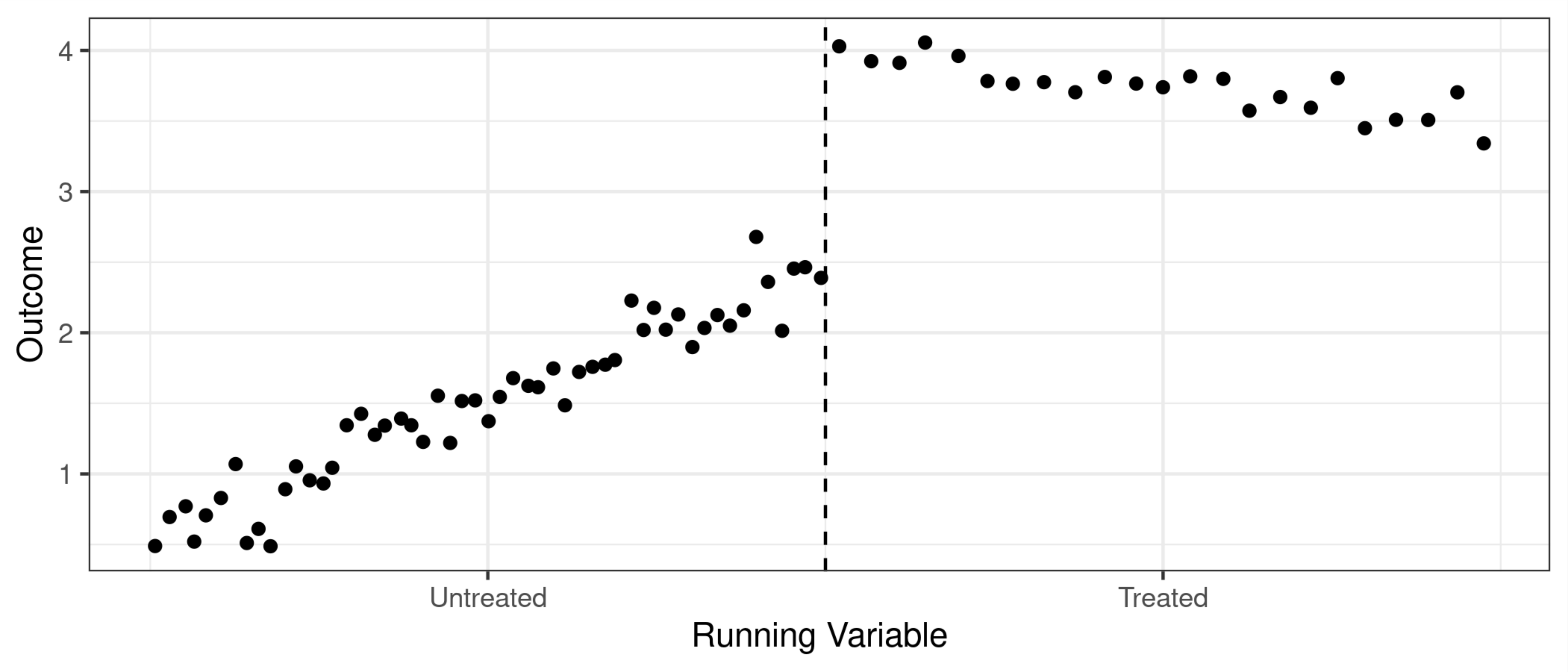
The binned averages helps to remove noise in the graph and can provide a cleaner look at the data. Just make sure that no bin includes observations above and below the cutoff!

Binned average calculation

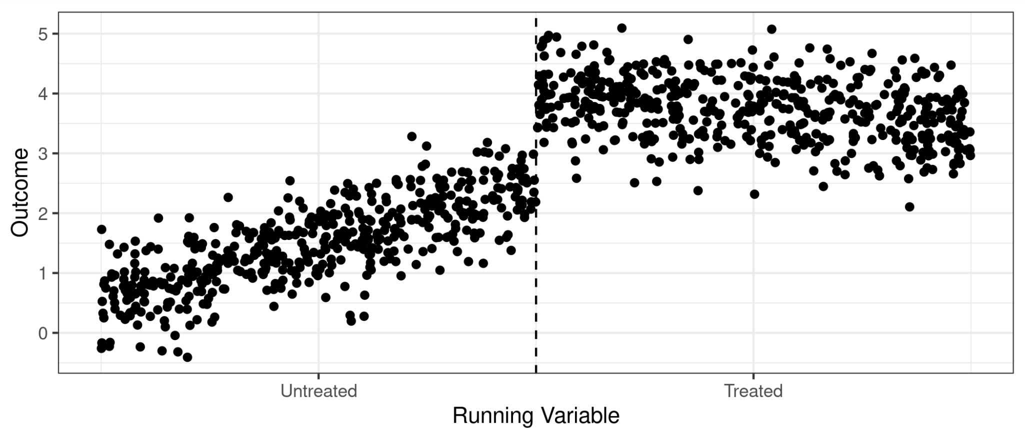
```r
library(rdrobust)
rd.result <- rdplot(rd.dat$Y, rd.dat$X,
                    c=1,
                    title="RD Plot with Binned Average",
                    x.label="Running Variable",
                    y.label="Outcome")
```

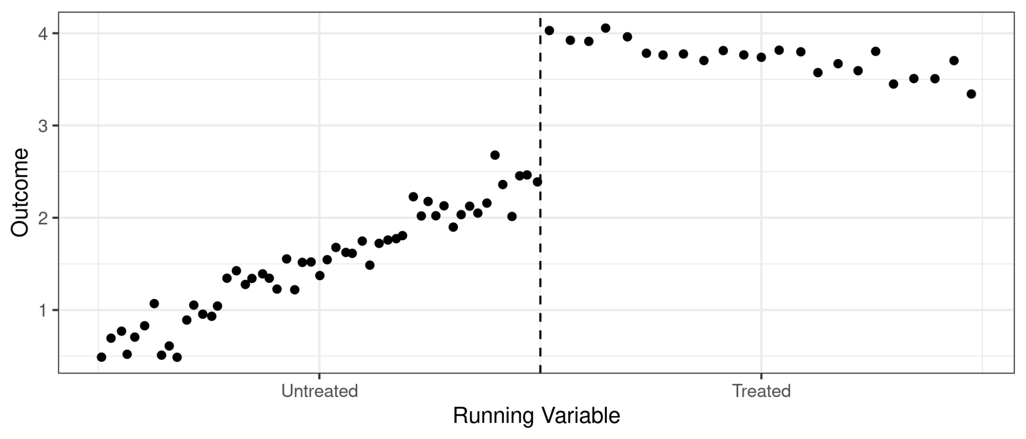
```r
bin.avg <- as_tibble(rd.result$vars_bins)

plot.bin <- bin.avg %>% ggplot(aes(x=rdplot_mean_x,y=rdplot_mean_y)) +
  geom_point() + theme_bw() +
  geom_vline(aes(xintercept=1),linetype='dashed') +
  scale_x_continuous(
    breaks = c(.5, 1.5),
    label = c("Untreated", "Treated")
  ) +
  xlab("Running Variable") + ylab("Outcome")
```

Binned average plot



With and without binning





Selecting "bin" width

1. Dummy variables: Create dummies for each bin, regress the outcome on the set of all dummies and form r-square 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.

2. Balance

Assessing balance

- If RD is an appropriate design, passing the cutoff should **only** affect treatment and outcome of interest
- How do we test for this?
  - Covariate balance
  - Placebo tests of other outcomes (e.g., t-1 outcomes against treatment at time t)

3. Manipulation tests

Manipulation of running variable

- Individuals should not be able to **precisely** manipulate running variable to enter into treatment
- Sometimes discussed as "bunching"
- Test for differences in density to left and right of cutoffs (`rddensity` in Stata and R)
- Permutation tests proposed in Ganong and Jager (2017)

What if bunching exists?

- Gerard, Rokkanen, and Rothe (2020) suggest partial identification allowing for bunching
- Can also be used as a robustness check
- `rdbounds` in Stata and R
- Assumption: bunching only moves people in one direction

4. RD Estimation

Baseline RD estimates

Start with the "default" options

- Local linear regression
- Optimal bandwidth
- Uniform kernel

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

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

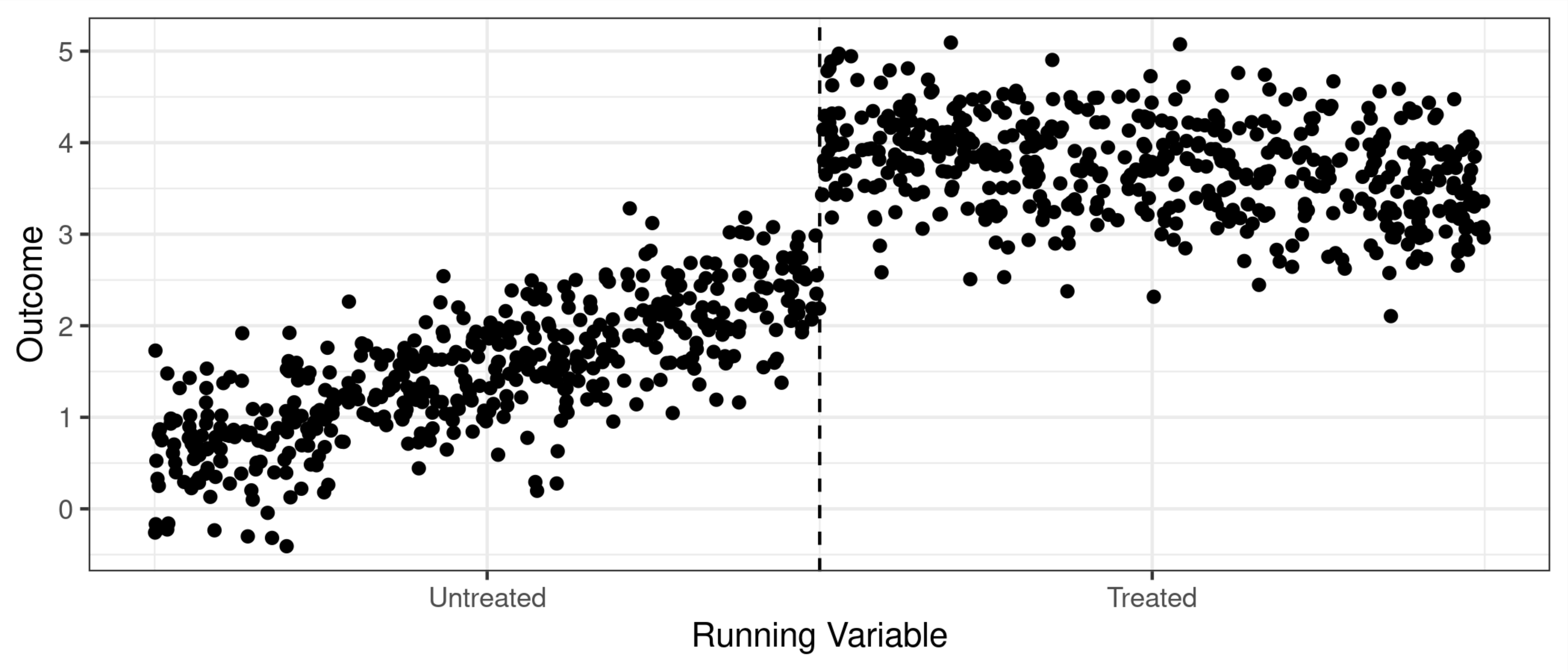
Cross-validation

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

```
ols ← lm(Y~X+W, data=rd.dat)

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

- True effect: 1.5
- Standard linear regression with same slopes: 1.68
- RD (linear with same slopes): 1.58

RD with built-in commands

```
## Sharp RD estimates using local polynomial regression.
##
## Number of Obs.                 1000
## BW type                       mserd
## Kernel                   Triangular
## VCE method                       NN
##
## Number of Obs.                  482          518
## Eff. Number of Obs.             146          187
## Order est. (p)                    1            1
## Order bias  (q)                   2            2
## BW est. (h)                   0.330        0.330
## BW bias (b)                   0.476        0.476
## rho (h/b)                     0.693        0.693
## Unique Obs.                     482          518
##
## =============================================================================
##         Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## =============================================================================
##   Conventional     1.593     0.108    14.732     0.000     [1.381 , 1.805]
##         Robust         -         -    12.530     0.000     [1.351 , 1.852]
## =============================================================================
```

Cattaneo et al. (2020) argue:

- Report conventional point estimate
- Report robust confidence interval

5. Robustness and sensitivity

Other options

- Different bandwidths
- Different kernels or polynomials
- Role of covariates in RD estimates

Pitfalls of polynomials

- Assign too much weight to points away from the cutoff
- Results **highly** sensitive to degree of polynomial
- Narrow confidence intervals (over-rejection of the null)

For more discussion, see this World Bank Blog post

Fuzzy RD

The Idea

"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,
  $E[Y_i|D_i = 1] - E[Y_i|D_i = 0] = \frac{E[Y_i|x_i \geq c] - E[Y_i|x_i < c]}{E[D_i|x_i \geq c] - E[D_i|x_i < c]}$
- Indicator for $x_i \geq c$ is an instrument for treatment status, D_i.
- Implemented with `rdrobust` and `fuzzy=t` option