NEBRASKA ACADEMY FOR
METHODOLOGY, ANALYTICS & PSYCHOMETRICS

Regression Discontinuity Designs in Social Science Research: Causal Inference of Cutoff-based Programs

HyeonJin Yoon, Ph.D.

3/13/2020

Nebraska Methodology Applications

University of Nebraska Lincoln

Agenda

- Background
- Rationale & Utility
- RDD Modeling and Analysis
  - Assumptions
  - Graphical, parametric, and nonparametric analyses
- Demonstration
- RDD Limitations
- RDD Extensions
- Conclusion

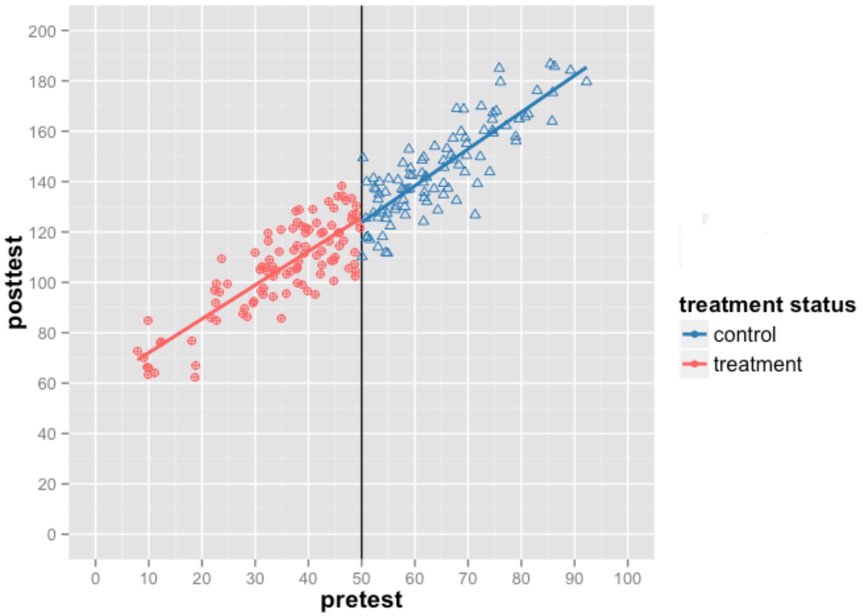
Causal inference in Social Science Research

- Causal inference of programs, interventions, and policies
  - Interested in answering questions if programs and policies are "effective" in improving outcomes
  - "Can subsidized employment programs help disadvantaged job seekers?"
  - "Can after school programs improve student social- and emotional, behavioral, and academic outcomes?"
- Randomized controlled trials (RCT)—"Gold Standard"
  - Could be neither feasible nor ethical to implement in practice

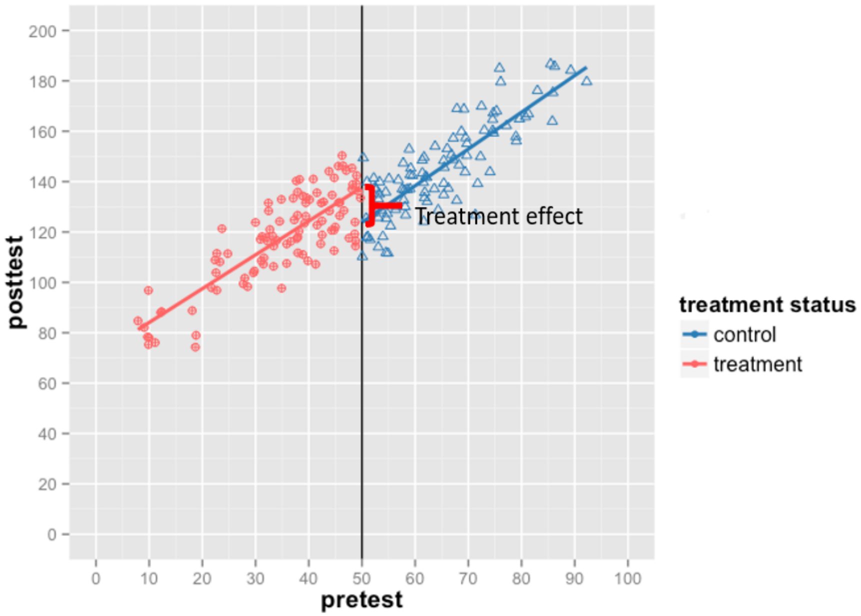
Regression Discontinuity Designs (RDD)

- A strong alternative to the RCT where a cutoff-based assignment of individuals is used
- Subjects are assigned to either the treatment or control condition based on a cutoff score on an assignment variable
  - Summer school reading programs
  - US minimum legal age of for drinking alcohol is 21
- When treatment is effective, a discontinuity in the regression relationship between assignment variable and outcome variable occurs at the cutoff
  - $Outcome_i = b_0 + b_1\,(assignment\ score_i) + b_2\,(\text{treatment}_i) + r_i$

Regression Discontinuity Designs (RDD)



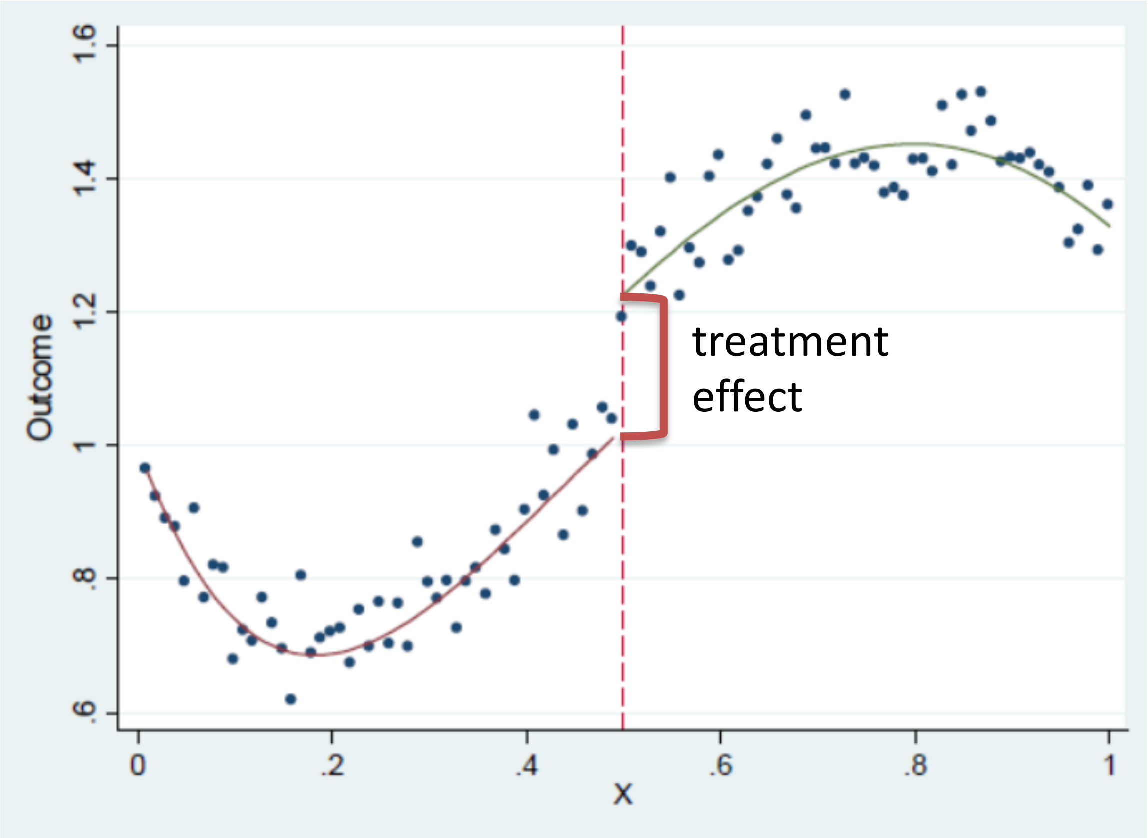
RD with no treatment effect



RD with treatment effect

MAP ACADEMY

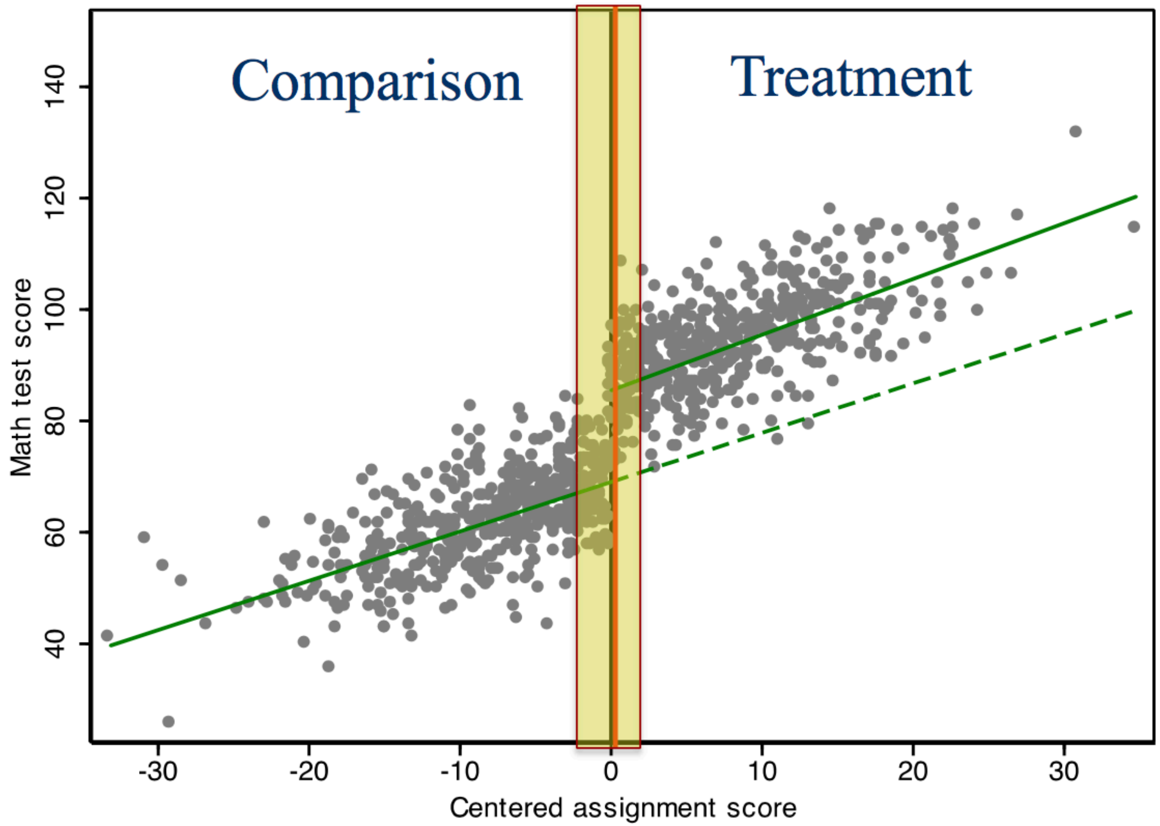
Regression Discontinuity Designs (RDD)



Causal Inference in RDD

- Treatment assignment is completely known and statistically modeled
  - Selection bias is controlled
- Local randomization at the cutoff
  - Participants just above and below the cutoff are assumed to be *identical*, except in terms of the treatment assignment
  - RDD assumptions are critical to prove this
- Produces unbiased causal estimate **at the cutoff**

Causal Inference in RDD



Causal Estimand of Interest in RDD

The treatment effect is the difference in the potential outcomes at the cutoff:

$$\tau_{ATE(C)} = E[Y_i(1) - Y_i(0) \mid Z_i = z_c]$$

$$= E[Y_i(1) \mid Z_i = z_z] - E[Y_i(0) \mid Z_i = z_c]$$

Advantage & Utility of RDD

- Allows the estimation of ***unbiased causal estimates*** at the cutoff in the design
  - due to the completely known selection process and local randomization occurring at the cutoff)
  - RDD causal estimates are as robust as those from RCT
- Enables program administrators to target those who are most in need of treatment
- Widely used in social science research evaluating a non-random, cut-off based programs and policies

RDD Assumptions

RDD Assumptions

1. Treatment assignment is measured and determined by a clearly known rule
   - Unit i is assignment to treatment condition if the unit scores below the cutoff $(Z_i=1 \text{ if } X_i < c)$, condition
   - Unit i is assignment to control condition if the unit scores above the cutoff $(Z_i=0 \text{ if } X_i \geq c)$
2. No alternative explanations for the treatment effect except through the treatment at the cutoff
   - Evidence of local randomization at the cutoff
   - Covariate balance at the cutoff, continuity of density at the cutoff (no manipulation of the treatment assignment)

RDD Modeling and Analysis

Types of RDD

- Sharp RDD: No non-compliance
- Fuzzy RDD: Non-compliance (no-shows, cross-overs)

Steps for RDD Modeling and Analysis

1. Assumption tests
2. Parametric analysis
3. Non-parametric analysis
4. Graphical analyses are utilized in steps 1-3
5. Examine the results from multiple analyses all together (steps 1-4)

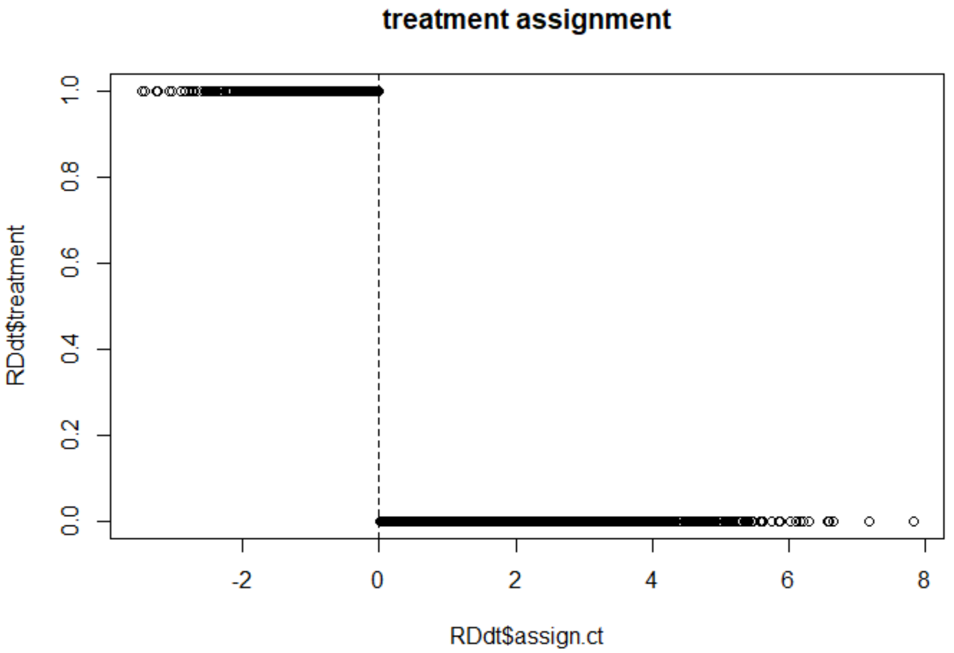
Components of RDD

- What is needed?

Components	Example
Assignment variable	Poverty score (composite)
Outcome variable	State test score
Condition variable	After school program (Yes or No)
Cut-score	25/100

Step 1: Assumption Test #1

- **Assumption # 1**: Treatment assignment is measured and determined by a clearly known rule
- Conditional probability of receiving the treatment jumps from 0 to 1 (or vise versa) at the cutoff



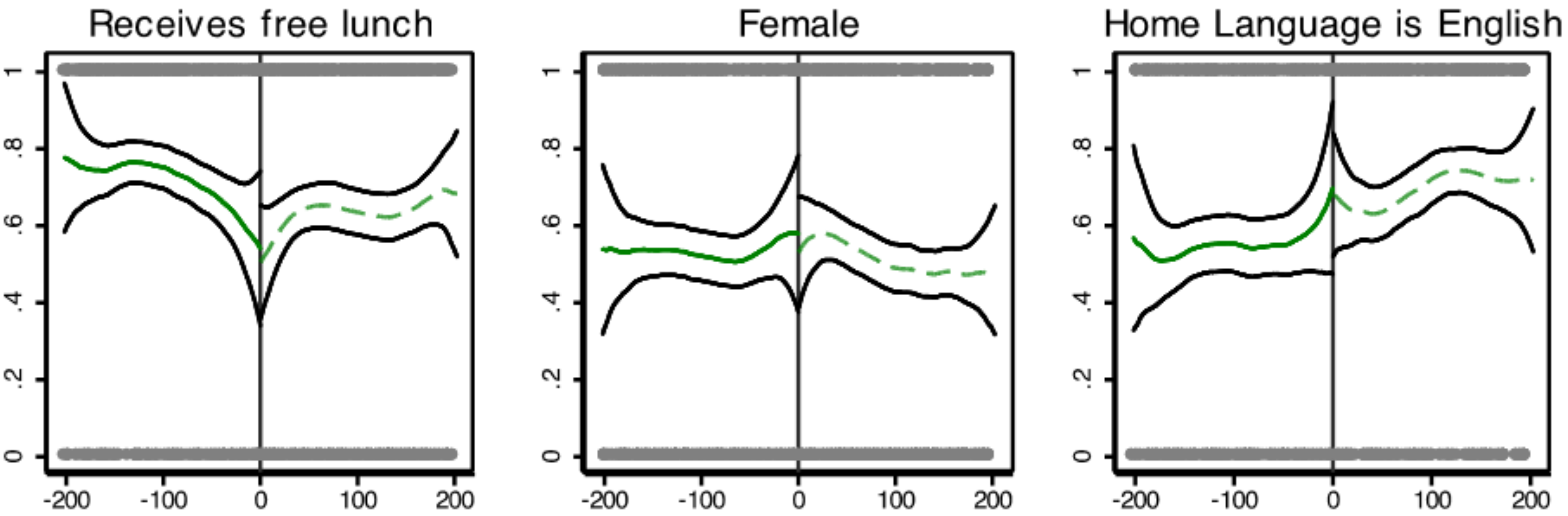
Step 1: Assumption Test #2

- **Assumption # 2:** No alternative explanations for the treatment effect except through the treatment at the cutoff
- **Covariate balance test**
  - No discontinuity in the potential outcomes (i.e., covariates)
  - Run a series of RD regressions with the baseline covariates
  - $Cov_i = \beta_0 + \beta_1 Treatment_i + \beta_2 f(Assign)_i + \beta_3 Treatment_i \times f(Assign)_i + \epsilon_i$
  - Create a series of scatterplots for the baseline covariates using nonparametric regression to model the relationship between the assignment variable and the outcome.

MAP ACADEMY

Step 1: Assumption Test #2

- Covariate balance plots



Step 1: Assumption Tests #2

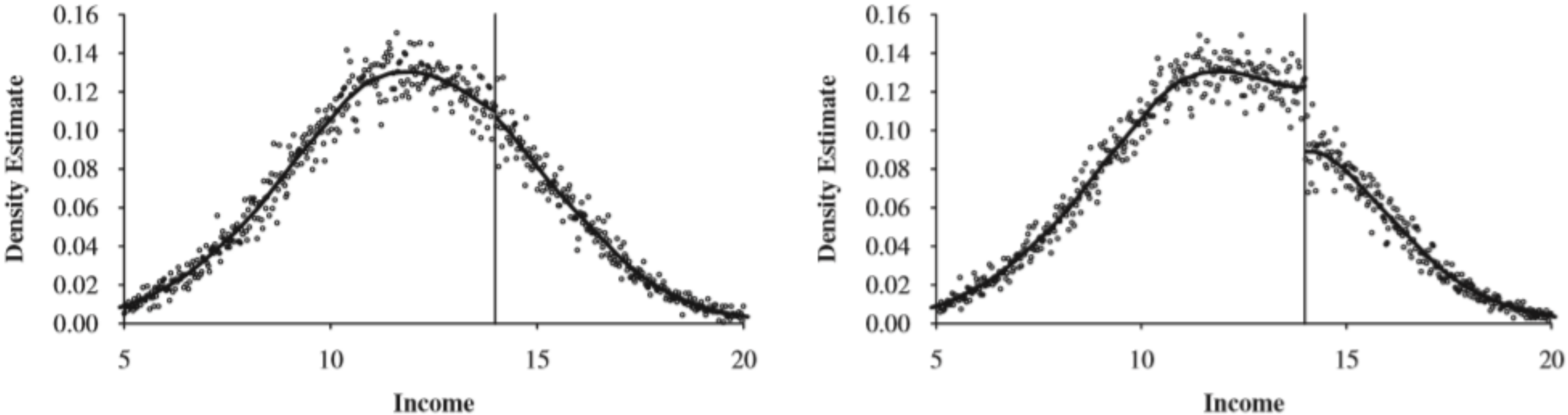
- **Assumption # 2:** No alternative explanations for the treatment effect except through the treatment at the cutoff
- **Density test (aka. Sorting test)**
  - Evaluates if there is manipulation of the assignment process
  - Assesses significant difference in number of observations at the cutoff
  - No discontinuity in the density of assignment variable at the cutoff
  - McCrary's Density Test (2008)

$$\theta = \ln \lim_{r \downarrow c} f(r) - \ln \lim_{r \uparrow c} f(r) \equiv \ln f^{+} - \ln f^{-}.$$

  - θ = log-difference in the height of each density function

Step 1: Assumption Test #2

- Density test example



Step 2: Parametric Analysis of the RDD

- A simple linear regression

$$Y_i = \alpha + \beta_0 \cdot T_i + \beta_1 \cdot r_i + \varepsilon_i$$

T = Treatment indicator,
r = Assignment variable

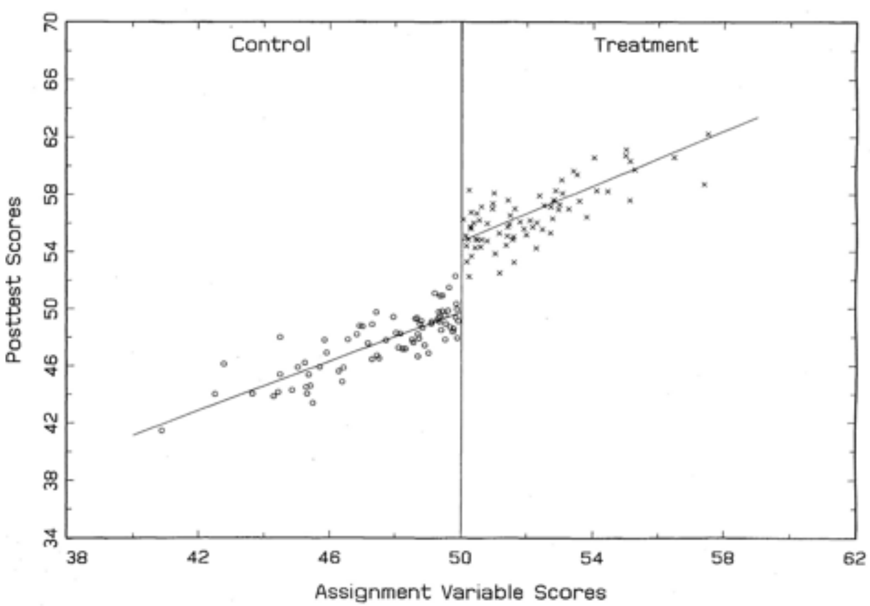
- Identify the best fitting model: *F*-test, AIC, BIC, LRT

1. linear $Y_i = \alpha + \beta_0 \cdot T_i + \beta_1 \cdot r_i + \varepsilon_i$
2. linear interaction $Y_i = \alpha + \beta_0 \cdot T_i + \beta_1 \cdot r_i + \beta_2 \cdot r_i \cdot T_i + \varepsilon_i$
3. quadratic $Y_i = \alpha + \beta_0 \cdot T_i + \beta_1 \cdot r_i + \beta_2 \cdot r_i^2 + \varepsilon_i$
4. quadratic interaction $Y_i = \alpha + \beta_0 \cdot T_i + \beta_1 \cdot r_i + \beta_2 \cdot r_i^2 + \beta_3 \cdot r_i \cdot T_i + \beta_4 \cdot r_i^2 \cdot T_i + \varepsilon_i$
5. cubic $Y_i = \alpha + \beta_0 \cdot T_i + \beta_1 \cdot r_i + \beta_2 \cdot r_i^2 + \beta_3 \cdot r_i^3 + \varepsilon_i$
6. cubic interaction $Y_i = \alpha + \beta_0 \cdot T_i + \beta_1 \cdot r_i + \beta_2 \cdot r_i^2 + \beta_3 \cdot r_i^3 + \beta_4 \cdot r_i \cdot T_i + \beta_5 \cdot r_i^2 \cdot T_i + \beta_6 \cdot r_i^3 \cdot T_i + \varepsilon_i$

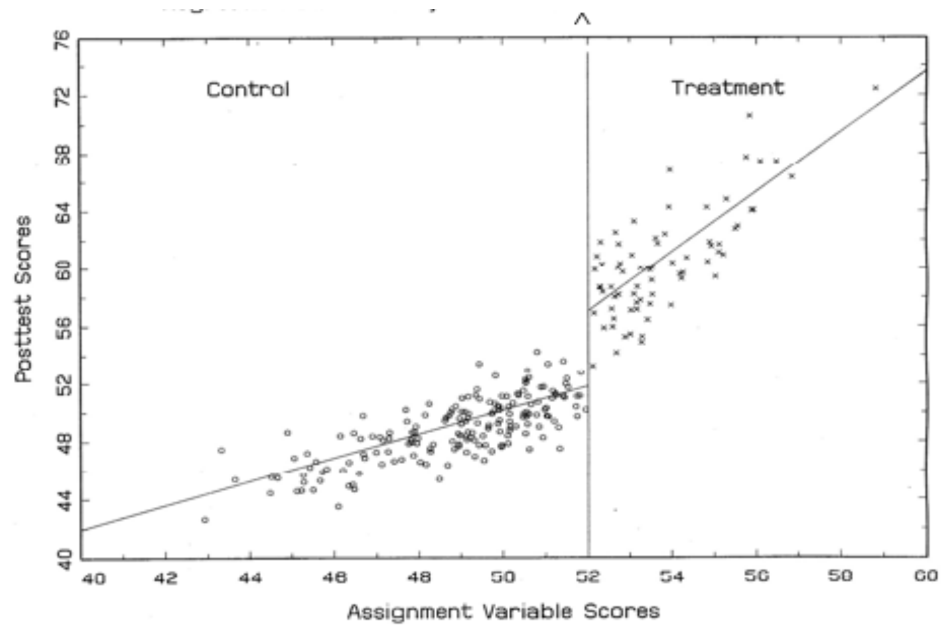
- Specification of the correct functional form is the key

Step 2: Parametric Analysis of the RDD

- Specification of the correct functional form is the key



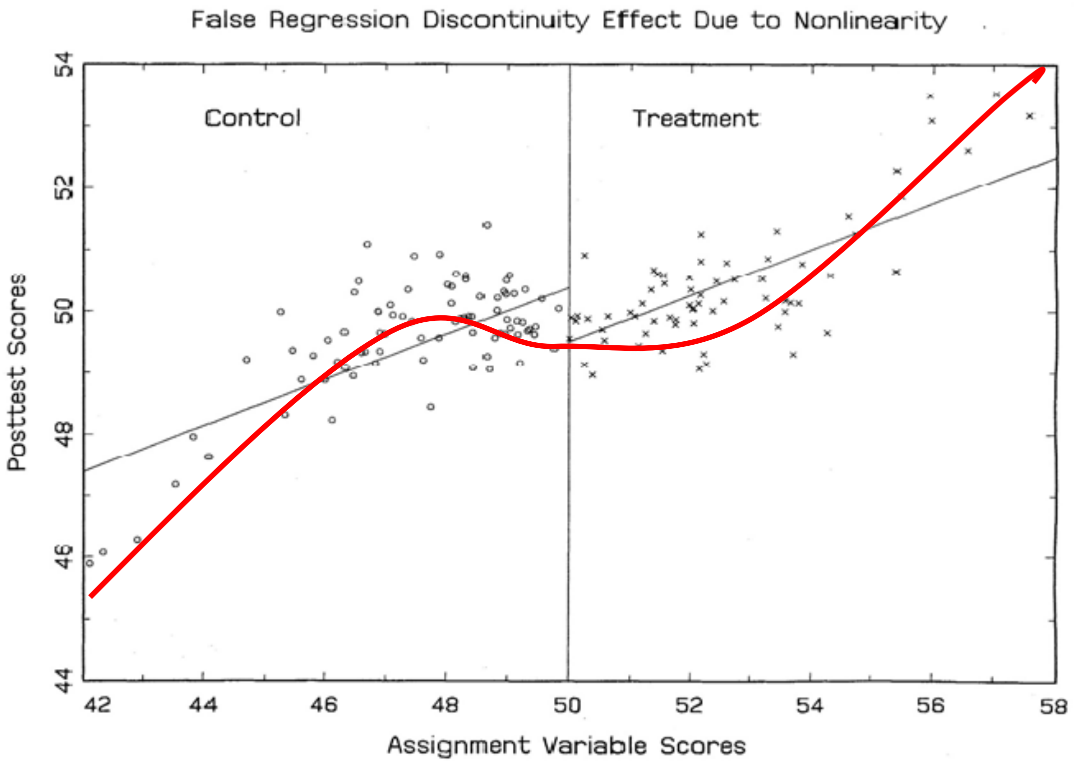
Constant treatment effect



Treatment x assignment score interaction

Step 2: Parametric Analysis of the RDD

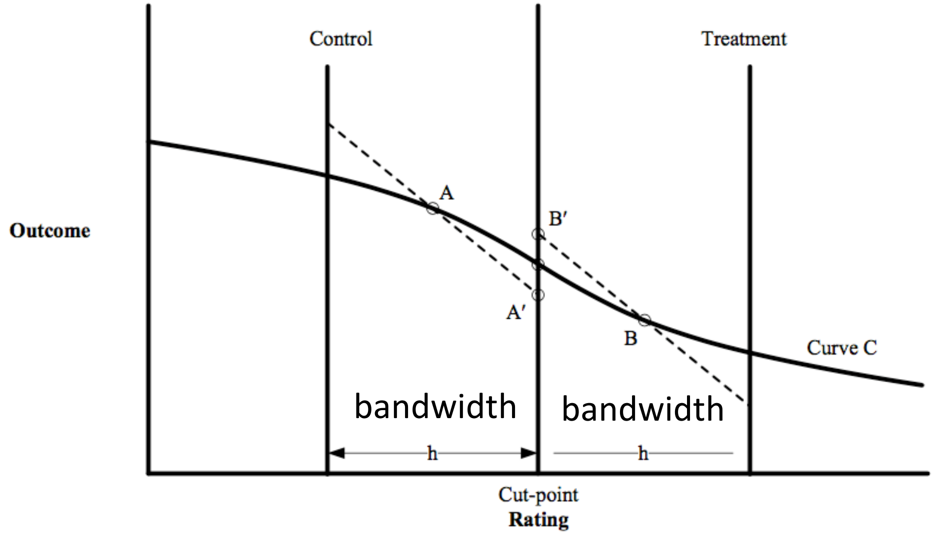
- Specification of the correct functional form is the key



MAP ACADEMY

Step 3: Nonparametric Analysis of the RDD

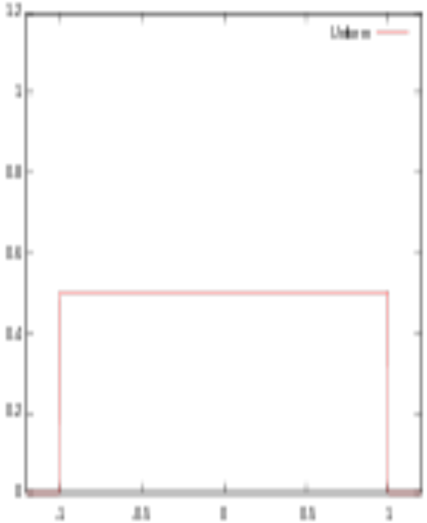
- Flexible for accommodating nonlinear relationship
- Local linear/polynomial regression is most commonly used in the literature
  1) Identify "**optimal bandwidth**"—a width of window where regressions are fitted



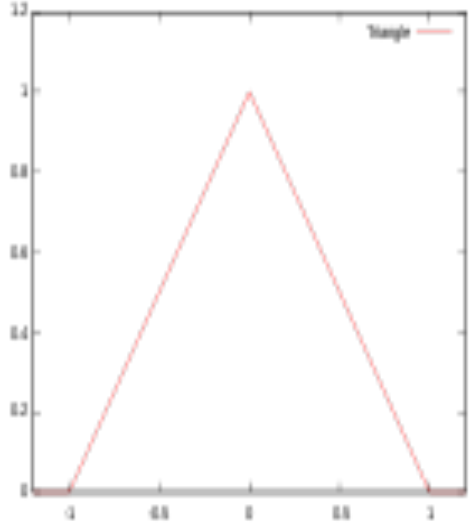
Step 3: Nonparametric Analysis of the RDD

- Local linear/polynomial regression

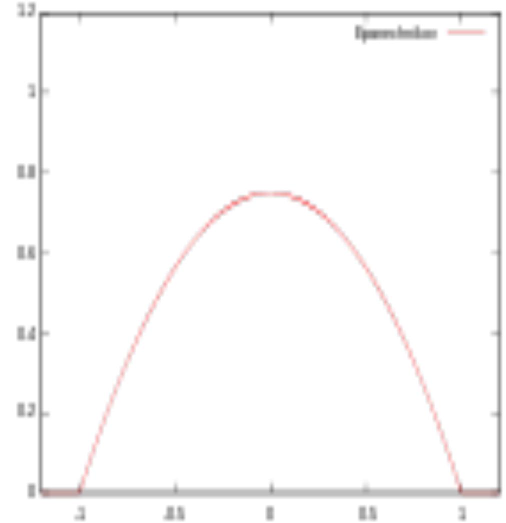
  2) Determine "**<u>kernel weights</u>**"



uniform



<u>triangle</u>



Epanechnikov

Step 3: Nonparametric Analysis of the RDD

- Local linear/polynomial regression

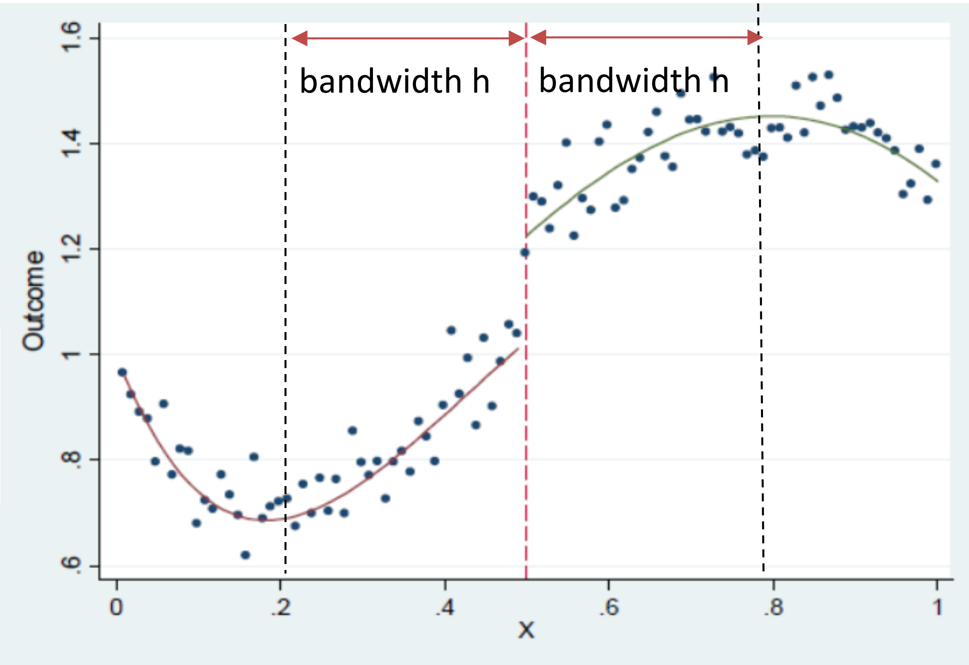
3) Regressions are fitted separately for observations on the left or right side of the cutoff.

$$Y_l = \alpha_l + \beta_l (X - c) + \varepsilon,$$

$$\text{where } c - h \leq X < c,$$

$$Y_r = \alpha_r + \beta_r (X - c) + \varepsilon,$$

$$\text{where } c \leq X \leq c + h.$$



Step 5: Convergence of Results

- Do the results from graphical, parametric, and nonparametric analyses converge in terms of direction and magnitude of the RD treatment effect?

Demonstration

RDD Analytic Software Packages

- **R**: rdrobust, rddensity, rdplot, rdd, RDDtools, mgcv (gam), rddapp
- **Stata**: lpoly (local polynomial regression), rdrobust,rddensity, rdplot

Dataset

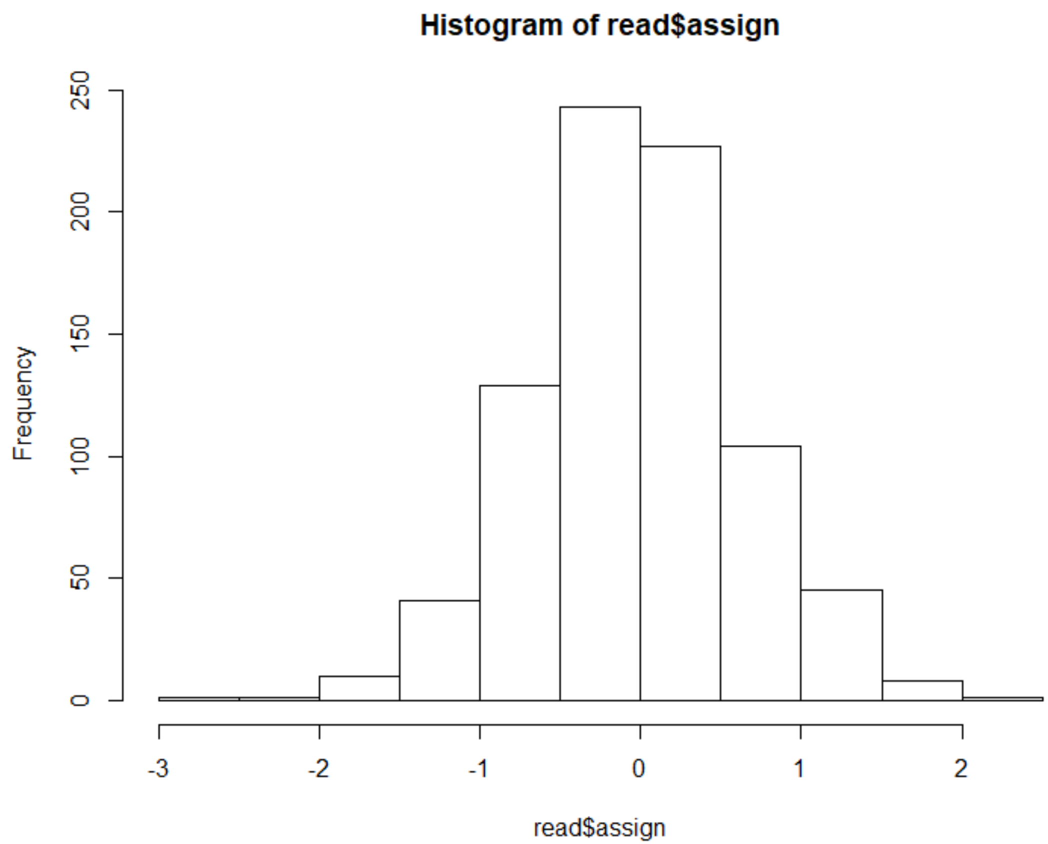
- Simulated dataset
- Treatment: Summer reading program
- Assignment variable: Composite standard score on a CBM reading assessment
- Outcome: State English Language Arts test score
- Cutoff = 0
- Sample: N = 810
  - (Treatment n = 425, Control n = 385)
- R package: rdrobust (Calonico et al., 2015)

Step 0: Explore the Data

- Summary statistics

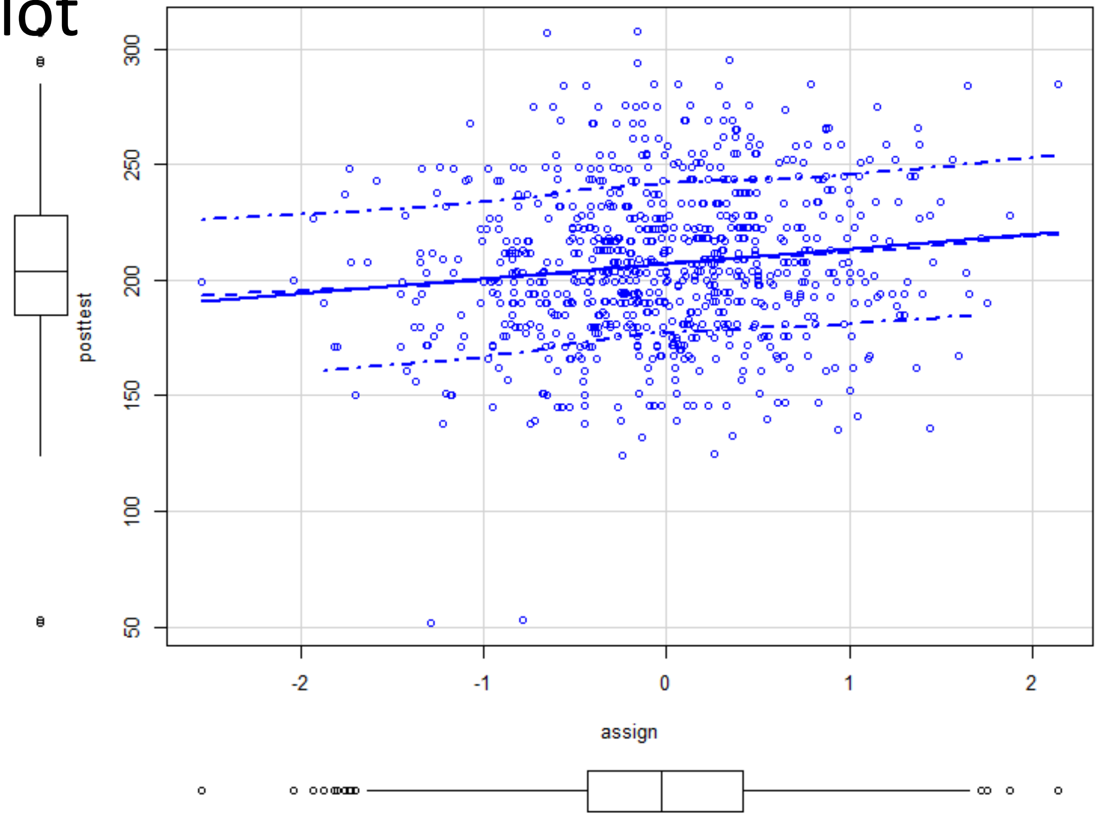
```
> # ::::: summary statistics :::::
> library(dplyr)
> read%>%
+   select(posttest, assign, treatment)%>%
+   describe()
         vars   n   mean     sd median trimmed    mad   min    max  range
posttest    1 810 207.13  33.13 204.00  206.83  34.10 52.00 308.00 256.00
assign      2 810  -0.02   0.67  -0.03   -0.02   0.63 -2.54   2.14   4.68
stuid       3 810 405.50 233.97 405.50  405.50 300.23  1.00 810.00 809.00
```

Step 0: Explore the Data



Step 0: Explore the Data

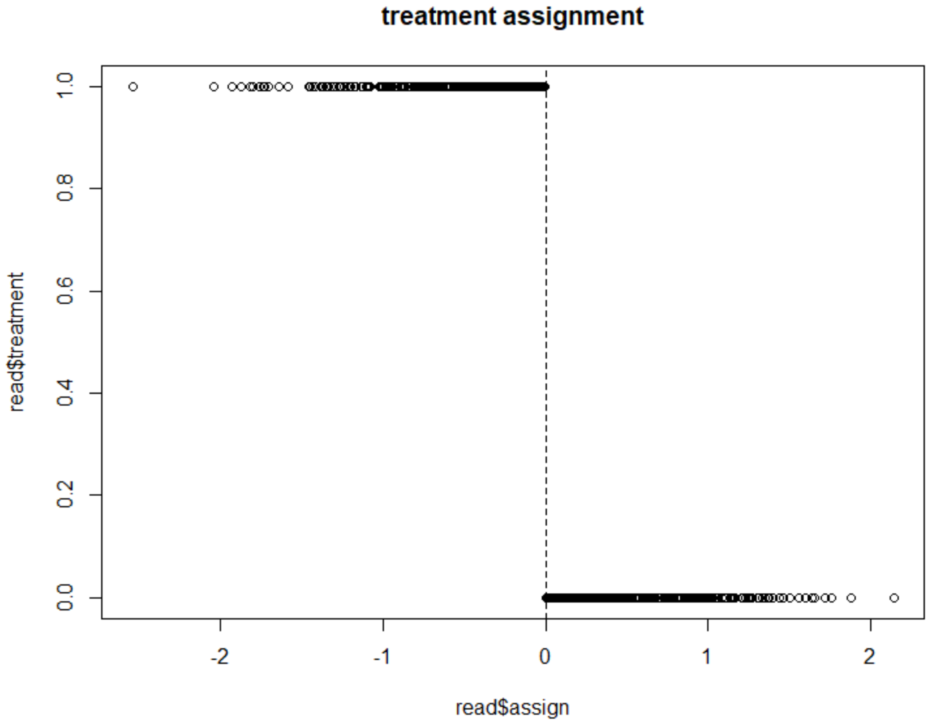
- Scatter plot



Step 1: Assumption Test

- Assignment Rule

```
# ::::: check the assignment
variable and assignment rule :::::
plot(read$assign, read$treatment,
main = "treatment assignment")
> abline(v = 0, lty = 2)
```



Step 3: Assumption Test-Covariate Balance

```
> summary(rdrobust(read$lep, read$assign, all=TRUE))
Call: rdrobust

Number of Obs.                    810
BW type                         mserd
Kernel                     Triangular
VCE method                         NN

Number of Obs.                   425          385
Eff. Number of Obs.              186          159
Order est. (p)                     1            1
Order bias  (q)                    2            2
BW est. (h)                    0.361        0.361
BW bias (b)                    0.638        0.638
rho (h/b)                      0.566        0.566
Unique Obs.                      422          382

=============================================================================
        Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
=============================================================================
  Conventional    -0.056     0.095    -0.586     0.558    [-0.242 , 0.131]
Bias-Corrected    -0.081     0.095    -0.855     0.392    [-0.268 , 0.105]
        Robust    -0.081     0.112    -0.726     0.468    [-0.301 , 0.138]
=============================================================================
```

Step 3: Assumption Test-Covariate Balance


```
> summary(rdrobust(read$sped, read$assign, all=TRUE))
Call: rdrobust

Number of Obs.                      810
BW type                           mserd
Kernel                       Triangular
VCE method                           NN

Number of Obs.                     425           385
Eff. Number of Obs.                226           197
Order est. (p)                       1             1
Order bias  (q)                      2             2
BW est. (h)                      0.449         0.449
BW bias (b)                      0.736         0.736
rho (h/b)                        0.610         0.610
Unique Obs.                        422           382

=============================================================================
        Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
=============================================================================
  Conventional     0.092     0.057     1.600     0.110    [-0.021 , 0.205]
Bias-Corrected     0.096     0.057     1.665     0.096    [-0.017 , 0.208]
        Robust     0.096     0.067     1.421     0.155    [-0.036 , 0.228]
=============================================================================
```

Step 2: Assumption Test – Sorting Test

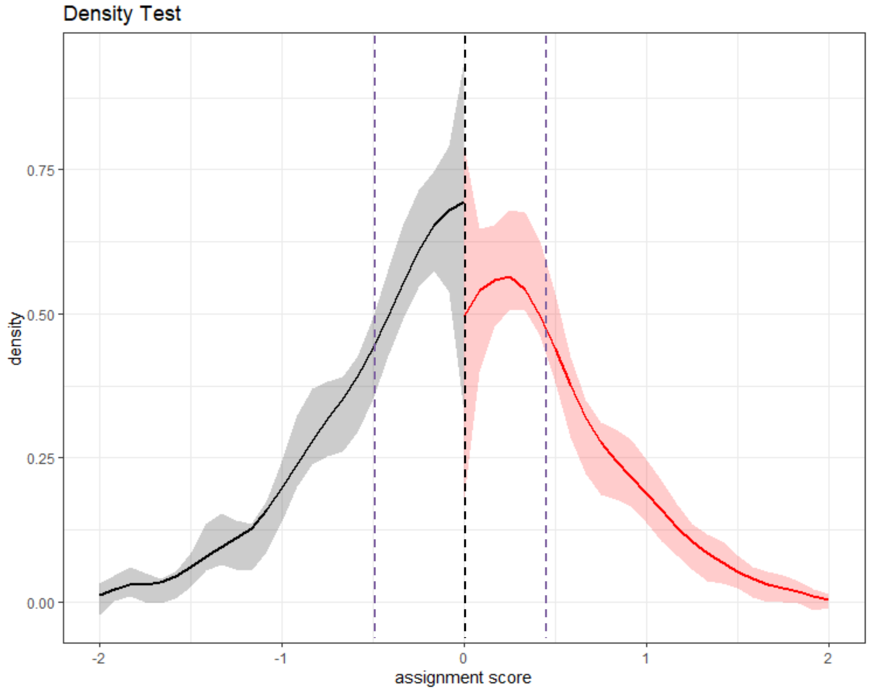
```
> #Sorting Around the Cutoff (McCrary Test)#
> summary(rddensity(X = read$assign))

RD Manipulation Test using local polynomial density estimation.

Number of obs =       810
Model =               unrestricted
Kernel =              triangular
BW method =           comb
VCE method =          jackknife

Cutoff c = 0          Left of c           Right of c
Number of obs         425                 385
Eff. Number of obs    229                 181
Order est. (p)        2                   2
Order bias (q)        3                   3
BW est. (h)           0.454               0.411

Method                T                   P > |T|
Robust                -0.6656             0.5057
```



Step 3: Parametric Analysis

```
# ::::: parametric analysis :::::
summary(out1.lm <- lm(posttest ~ treatment*assign, data = read))                # linear mode:
summary(out2.lm <- lm(posttest ~ treatment*(assign + I(assign^2)), data = read))  # quadratic
summary(out3.lm <- lm(posttest ~ treatment*(assign + I(assign^2) + I(assign^3)), data = read))
anova(out1.lm, out2.lm, out3.lm)
AIC(out1.lm, out2.lm, out3.lm)
```

```
> anova(out1.lm, out2.lm, out3.lm)
Analysis of Variance Table

Model 1: posttest ~ treatment * assign
Model 2: posttest ~ treatment * (assign + I(assign^2))
Model 3: posttest ~ treatment * (assign + I(assign^2) + I(assign^3))
  Res.Df    RSS Df Sum of Sq      F  Pr(>F)
1    806 871233
2    804 870011  2    1222.7 0.5682 0.56675
3    802 862818  2    7192.8 3.3429 0.03583 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> AIC(out1.lm, out2.lm, out3.lm)
        df      AIC
out1.lm  5 7962.991
out2.lm  7 7965.854
out3.lm  9 7963.129
```

Step 3: Parametric Analysis

```
# ::::: parametric analysis :::::
summary(out1.lm <- lm(posttest ~ treatment*assign, data = read))              # linear mode:
summary(out2.lm <- lm(posttest ~ treatment*(assign + I(assign^2)), data = read))   # quadratic
summary(out3.lm <- lm(posttest ~ treatment*(assign + I(assign^2) + I(assign^3)), data = read))
anova(out1.lm, out2.lm, out3.lm)
AIC(out1.lm, out2.lm, out3.lm)
```

```
> anova(out1.lm, out2.lm, out3.lm)
Analysis of Variance Table

Model 1: posttest ~ treatment * assign
Model 2: posttest ~ treatment * (assign + I(assign^2))
Model 3: posttest ~ treatment * (assign + I(assign^2) + I(assign^3))
  Res.Df    RSS Df Sum of Sq      F  Pr(>F)
1    806 871233
2    804 870011  2    1222.7 0.5682 0.56675
3    802 862818  2    7192.8 3.3429 0.03583 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> AIC(out1.lm, out2.lm, out3.lm)
        df      AIC
out1.lm  5 7962.991
out2.lm  7 7965.854
out3.lm  9 7963.129
```

Step 3: Parametric Analysis

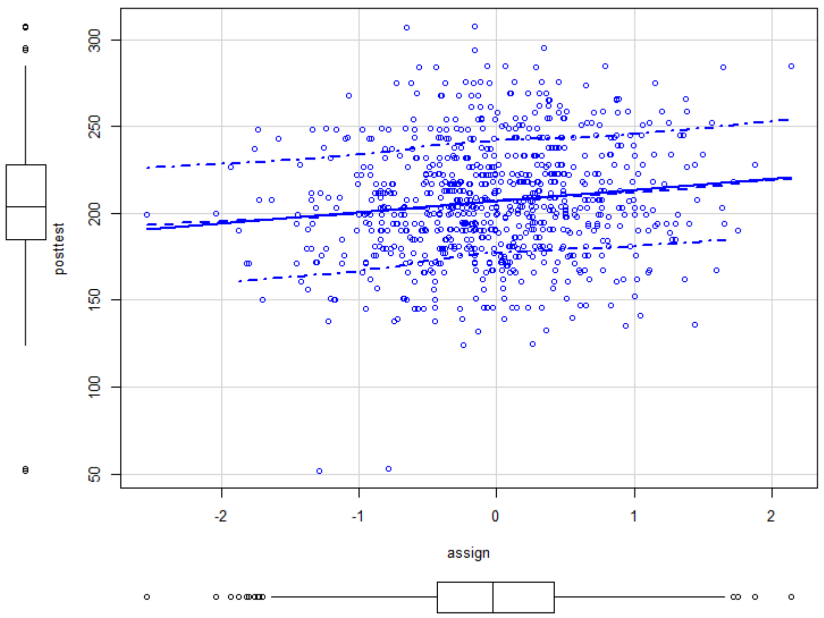
- Parametric Results

```
Call:
lm(formula = posttest ~ treatment * (assign + I(assign^2) + I(assign^3)),
    data = read)

Residuals:
    Min      1Q  Median      3Q     Max
-147.61  -20.95   -2.12   21.64  104.66

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)            197.936      5.135  38.544   <2e-16 ***
treatment               13.653      6.609   2.066   0.0392 *
assign                  64.840     26.688   2.430   0.0153 *
I(assign^2)            -88.490     36.273  -2.440   0.0149 *
I(assign^3)             34.842     13.659   2.551   0.0109 *
treatment:assign       -51.882     33.536  -1.547   0.1222
treatment:I(assign^2)   84.527     44.070   1.918   0.0555 .
treatment:I(assign^3)  -38.341     15.971  -2.401   0.0166 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 32.8 on 802 degrees of freedom
Multiple R-squared:  0.02849,	Adjusted R-squared:  0.02001
F-statistic:  3.36 on 7 and 802 DF,  p-value: 0.001541
```



Step 4: Nonparametric Analysis

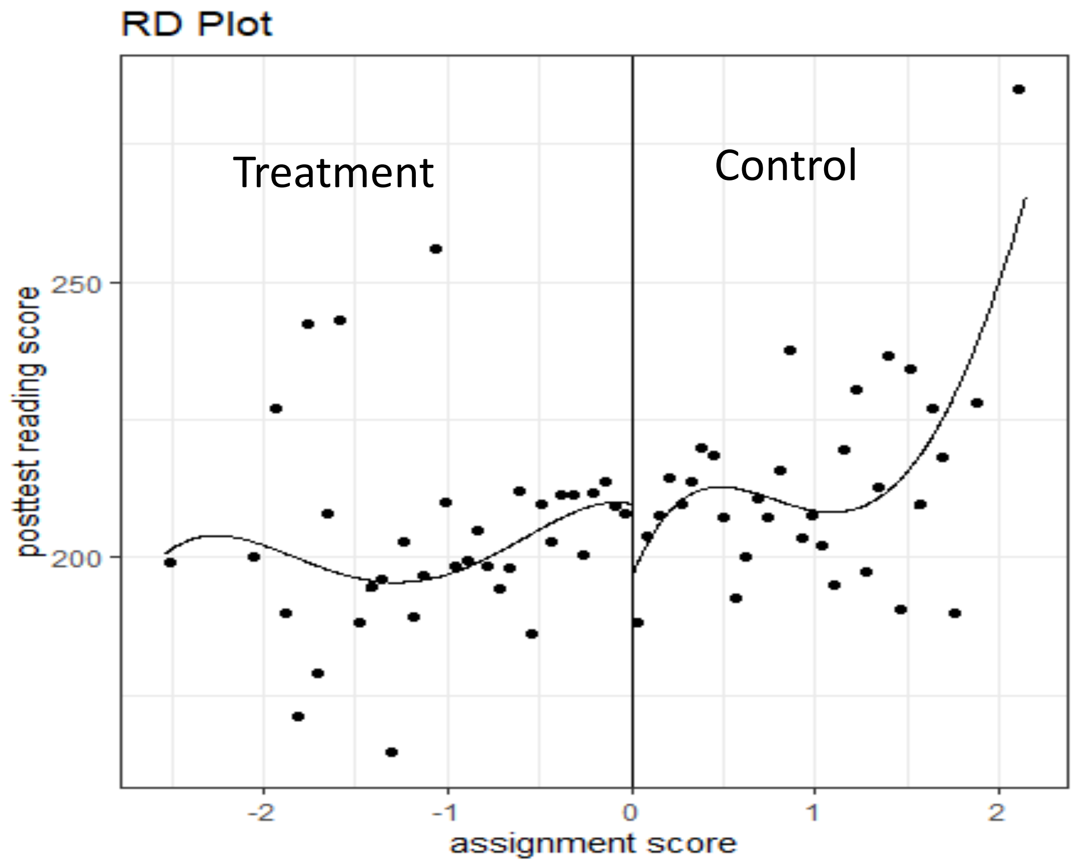
```
> # Local Polynomial Regression#
> summary(rdrobust(y = read$posttest, x = read$assign, all = TRUE))
Call: rdrobust

Number of Obs.                  810
BW type                       mserd
Kernel                   Triangular
VCE method                       NN

Number of Obs.                 425          385
Eff. Number of Obs.            194          164
Order est. (p)                   1            1
Order bias  (q)                  2            2
BW est. (h)                  0.374        0.374
BW bias (b)                  0.627        0.627
rho (h/b)                    0.595        0.595
Unique Obs.                    422          382

=============================================================================
        Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
=============================================================================
  Conventional   -14.469     7.244    -1.997     0.046   [-28.667 , -0.270]
Bias-Corrected   -16.639     7.244    -2.297     0.022   [-30.837 , -2.440]
        Robust   -16.639     8.441    -1.971     0.049   [-33.184 , -0.094]
```

Step 4: Nonparametric Analysis



Step 5: Cross-check the Results

Model	Treatment effect (SE)
Linear regression	4.61 (3.77)
Quadratic regression	6.33 (5.19)
Cubic regression*	13.65 (6.61)*
Local linear regression (conventional)	14.47 (7.24)*
Local linear regression (robust, bias-corrected)	16.65 (8.44)*

Limitations of RD

- **Lower statistical power than a comparable RCT**
  - Due to the correlation between the treatment status indicator and the assignment variable
  - RD requires a sample size between 2.75 and 4 times greater than that of a comparable RCT to detect the same treatment effects
- **Strong reliance on correct modeling of the assignment variable-outcome relationship**
  - If researchers modeled a linear function when the true function for the hypothesized relationship is not linear (e.g., curvilinear), they might find an artifactual discontinuity at the cutoff
- **Limited generality of causal inference**
  - Causal inference in basic RD is limited to the small area surrounding the cutoff

Conclusion

- RDD enables the ethical delivery of programs and policy in social science research.
- RDD yields unbiased causal estimate at the cutoff.
- Assumption tests are crucial.
- RDD requires a large sample.
- Large sample at the cutoff matters.
- Cross-check the results from different types of RD analyses (i.e., parametric, non-parametric, and graphical analyses).

RDD Extensions

- RDD variants to improve generality of RD estimate beyond the cutoff/also improve powers
  - Comparative RDD (using pretest scores or non-equivalent groups)
  - RDD with covariate matching
  - Multiple-cutoff RDD

Questions?

Thank you!

Fuzzy RDD

- Non-compliance!
- Wald Estimator approach

$$LATE_C = \frac{\lim\limits_{z \uparrow z_c} E[Y_i \mid Z_i = z_c] - \lim\limits_{z \downarrow z_c} E[Y_i \mid Z_i = z_c]}{\lim\limits_{z \uparrow z_c} E[D_i \mid Z_i = z_c] - \lim\limits_{z \downarrow z_c} E[D_i \mid Z_i = z_c]} .$$

- Two-state least squares (2SLS) approach
  - First stage: $Treated_i = \beta_0 + \beta_1 Treatment_i + \beta_2 g(AVAR)_i + \varepsilon_i$
  - Second stage: $Y_i = \beta_0 + \beta_1 \widehat{Treated}_i + \beta_2 f(AVAR)_i + u_i$