Methods: Mind the Gap

Webinar Series

Regression Discontinuity Designs

Presented by:

Matias D. Cattaneo, Ph.D.
Princeton University

National Institutes of Health
Office of Disease Prevention

Webinar on Regression Discontinuity Designs

Methods: Mind the Gap Webinar Series

Office of Disease Prevention, National Institutes of Health

Matias D. Cattaneo

Princeton University

https://cattaneo.princeton.edu/

January 17, 2024

Complementary materials available at https://rdpackages.github.io/

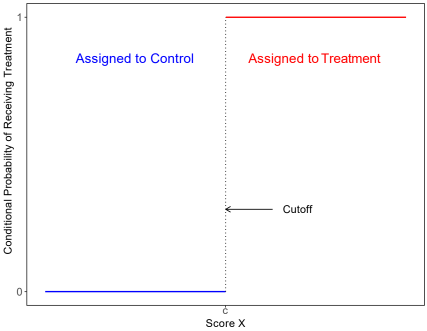
Outline

Causal Inference and Program Evaluation

- Main goal: learn about treatment effect of policy or intervention
- If treatment randomization available → easy to estimate effects
- If treatment randomization not available → observational studies
  - Selection on observables.
  - Instrumental variables, etc.
- **Regression discontinuity (RD) design**
  - Simple assignment, based on known external factors
  - Objective basis to evaluate assumptions
  - Easy to falsify and interpret.
  - *Careful*: very local!

Regression Discontinuity Design

- Units receive a **score** (X_i).
- A treatment is assigned based on the score and a *known* **cutoff** (c).
- The **treatment** is:
  - given to units whose score is greater than the cutoff.
  - withheld from units whose score is less than the cutoff.
- Under assumptions, the abrupt change in the probability of treatment assignment allows us to learn about the effect of the treatment.



RD Designs: Taxonomy

- **Frameworks**.
  - Identification: Continuity/Extrapolation, Local Randomization.
  - Score: Continuous, Many Repeated, Few Repeated.
- **Settings**.
  - Sharp, Fuzzy, Kink, Kink Fuzzy.
  - Multiple Cutoff, Multiple Scores, Geographic RD.
  - Dynamic, Continuous Treatments, Time, etc.
- **Parameters of Interest**.
  - Average Effects, Quantile/Distributional Effects, Partial Effects.
  - Heterogeneity, Covariate-Adjustment, Differences, Time.
  - Extrapolation.

RCTs vs. (Sharp) RD Designs

- **Notation**: $(Y_i(0), Y_i(1), X_i)$, $i = 1, 2, \ldots, n$.
- **Treatment**: $T_i \in \{0, 1\}$, T_i independent of $(Y_i(0), Y_i(1), X_i)$.
- **Data**: (Y_i, T_i, X_i), $i = 1, 2, \ldots, n$, with

$$Y_i = \begin{cases} Y_i(0) & \text{if } T_i = 0 \\ Y_i(1) & \text{if } T_i = 1 \end{cases}$$

- **Average Treatment Effect**:

$$\tau_{\mathsf{ATE}} = \mathsf{E}[Y_i(1) - Y_i(0)] = \mathsf{E}[Y_i | T = 1] - \mathsf{E}[Y_i | T = 0]$$

RCTs vs. (Sharp) RD Designs

- **Notation**: $(Y_i(0), Y_i(1), X_i)$, $i = 1, 2, \ldots, n$, X_i score.

- **Treatment**: $T_i \in \{0, 1\}$, $T_i = (X_i \geq c)$, c cutoff.

- **Data**: (Y_i, T_i, X_i), $i = 1, 2, \ldots, n$, with

$$Y_i = \begin{cases} Y_i(0) & \text{if } T_i = 0 \\ Y_i(1) & \text{if } T_i = 1 \end{cases}$$

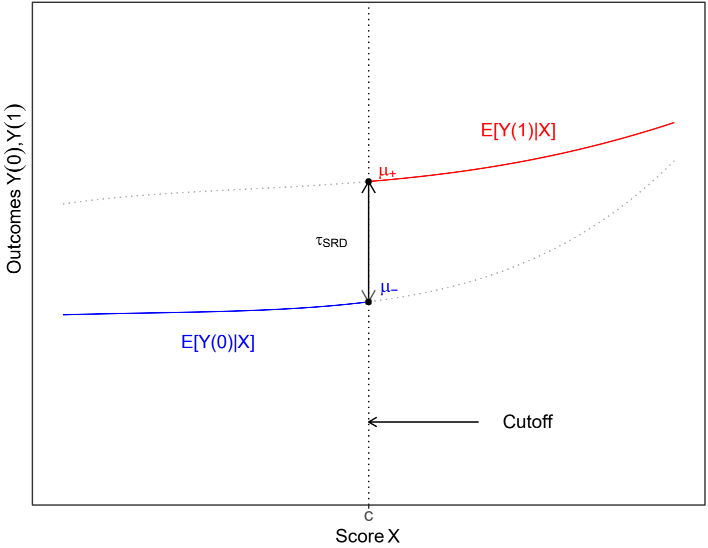
- **Average Treatment Effect at the cutoff** (Continuity-based):

$$\tau_{\mathsf{SRD}} = \mathsf{E}[Y_i(1) - Y_i(0) | X_i = c] = \lim_{x \downarrow c} \mathsf{E}[Y_i | X_i = x] - \lim_{x \uparrow c} \mathsf{E}[Y_i | X_i = x]$$

- **Average Treatment Effect in a neighborhood** (LR-based):

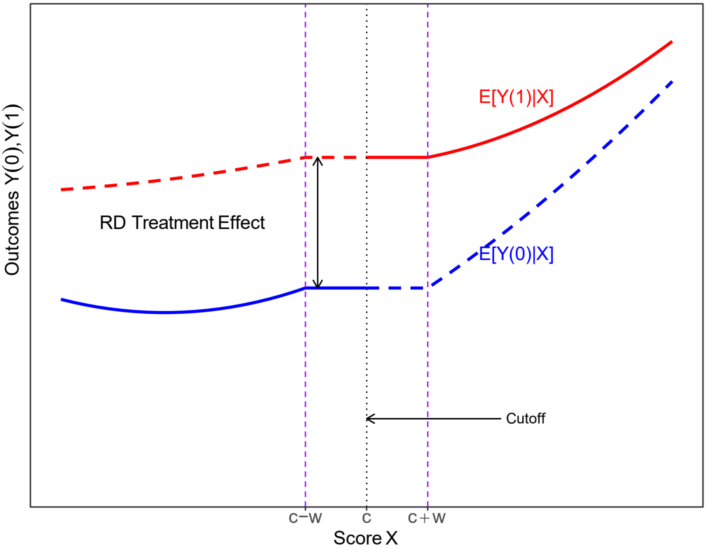
$$\tau_{\mathsf{LR}} = \frac{1}{N_{\mathsf{W}}} \sum_{X_i \in \mathsf{W}} \mathsf{E}[Y_i(1) - Y_i(0) | X_i \in \mathsf{W}] = \frac{1}{N_1} \sum_{X_i \in \mathsf{W}, T_i = 1} Y_i - \frac{1}{N_0} \sum_{X_i \in \mathsf{W}, T_i = 0} Y_i$$

$$\tau_{SRD} = \underbrace{E[Y_i(1) - Y_i(0) \mid X_i = c]}_{\text{Unobservable}} = \underbrace{\lim_{x \downarrow c} E[Y_i \mid X_i = x]}_{\text{Estimable}} - \underbrace{\lim_{x \uparrow c} E[Y_i \mid X_i = x]}_{\text{Estimable}}$$

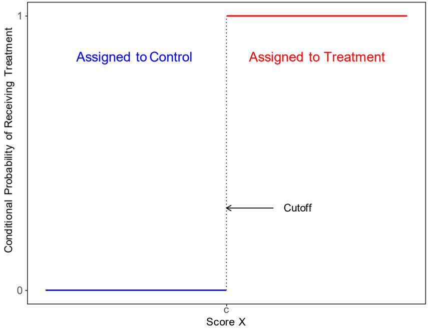


T_i independent of $(Y_i(0),\ Y_i(1))$ for all $X_i \in W = [c - w, c + w]$

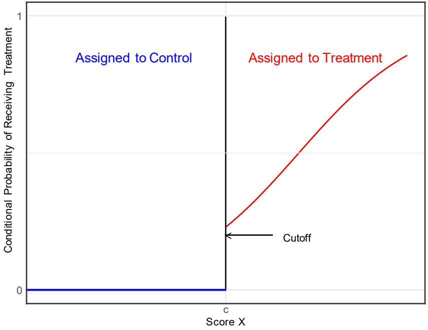
$+$ exclusion restriction



Fuzzy RD Designs



(a) Sharp RD



(b) Fuzzy RD (one-sided compliance)

Fuzzy RD Designs

- **Imperfect compliance.**
  - probability of receiving treatment changes at c, but not necessarily from 0 to 1.

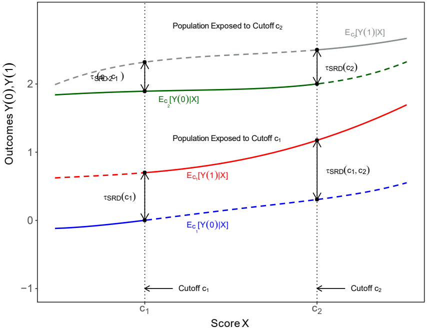
- Canonical Parameter:

$$\tau_{\mathsf{FRD}} = \frac{\mathsf{E}[(Y_i(1) - Y_i(0)(D_i(1) - D_i(0))|X_i = c]}{\mathsf{E}[D_i(1)|X_i = c] - \mathsf{E}[D_i(0)|X_i = c]}$$
$$= \frac{\lim_{x \downarrow c} \mathsf{E}[Y_i|X_i = x] - \lim_{x \uparrow c} \mathsf{E}[Y_i|X_i = x]}{\lim_{x \downarrow c} \mathsf{E}[D_i|X_i = x] - \lim_{x \uparrow c} \mathsf{E}[D_i|X_i = x]}$$

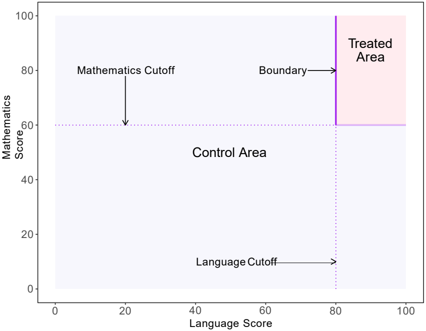
- Similarly for Local Randomization framework.

- Different interpretations under different assumptions.

Multi-cutoff, Multi-Score, Geographic RD Designs



(a) Multi-cutoff:
$\tau_{\mathsf{SRD}}(x, c) = \mathbb{E}[Y_i(1) - Y_i(0) | X_i = x, C_i = c]$



(b) Multi-score:
$\tau_{\mathsf{SRD}}(x_1, x_2) = \mathbb{E}[Y_i(1) - Y_i(0) | X_{1i} = x_1, X_{2i} = x]$

Highlights and Main Takeaways

- RD designs exploit "variation" near the cutoff.
- Causal effect is different (in general) than RCT.
- No "overlap" (sharp) so extrapolation or exclusion is unavoidable.
- Graphical analysis is both very useful and very dangerous.
- Need to work with data near cutoff $\Longrightarrow$ bandwidth or window selection.
- Many design-specific falsification/validation methods.

Outline

RD Packages: Python, R, Stata

https://rdpackages.github.io/

- **rdrobust**: estimation, inference and graphical presentation using local polynomials, partitioning, and spacings estimators.
  - rdrobust, rdbwselect, rdplot.
- **rddensity**: discontinuity in density tests (manipulation testing) using both local polynomials and binomial tests.
  - rddensity, rdbwdensity.
- **rdlocrand**: covariate balance, binomial tests, randomization inference methods (window selection & inference).
  - rdrandinf, rdwinselect, rdsensitivity, rdrbounds.
- **rdmulti**: multiple cutoffs and multiple scores.
- **rdpower**: power, sample selection and minimum detectable effect size.

Empirical Illustration: Head Start (Ludwig and Miller, 2007, QJE)

- **Problem**: impact of Head Start on Infant Mortality

- **Data**:

  $Y_i =$ child mortality 5 to 9 years old

  $T_i =$ whether county received Head Start assistance

  $X_i =$ 1960 poverty index ($c = 59.1984$)

  $Z_i =$ see database.

- **Potential outcomes**:

  $Y_i(0) =$ child mortality if **had not received** Head Start

  $Y_i(1) =$ child mortality if **had received** Head Start

- **Causal Inference**:

  $Y_i(0) \neq Y_i | T_i = 0$ and $Y_i(1) \neq Y_i | T_i = 1$

RD Plots

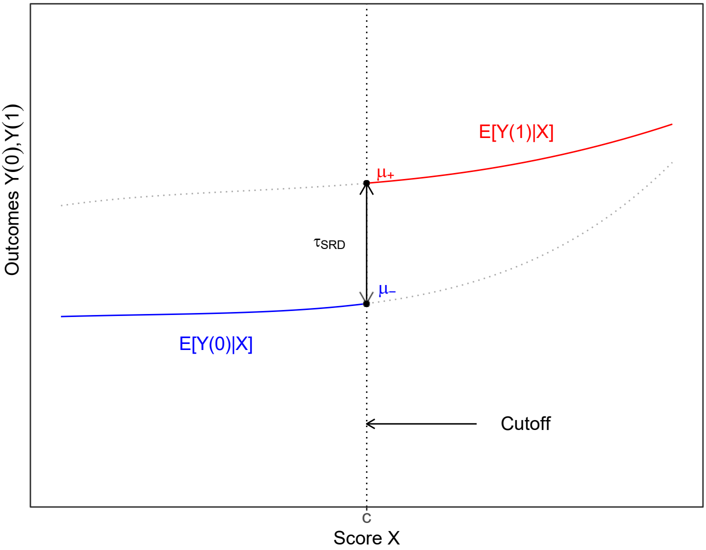
- Main ingredients:
  - Global smooth polynomial fit.
  - Binned discontinuous local-means fit.
- Main goals:
  - Graphical (heuristic) representation.
  - Detention of discontinuities.
  - Representation of variability.
- Tuning parameters:
  - Global polynomial degree.
  - Location (ES or QS) and number of bins.
- **Great to convey ideas but horrible to draw conclusions**.

Estimation and Inference Methods

- **Continuity/Extrapolation**: Local polynomial approach.
  - Localization: bandwidth selection (trade-off bias and variance).
  - Point estimation: "flexible" (nonparametric).
  - Inference: robust bias-corrected methods.
- **Local Randomization**: finite-sample and large-sample inference.
  - Localization: window selection (via local independence implications).
  - Point estimation: parametric, finite-sample (Fisher) or large-sample (Neyman/SP).
  - Inference: randomization inference (Fisher) or large-sample (Neyman/SP).
- Many refinements and other methods exist (EL, Bayesian, Uniformity, etc.).
  - Do not offer much improvements in applications.
  - Can be overly complicated (lack of transparency).
  - Can depend on user-chosen tuning parameters (lack of replicability).

Outline

$$\tau_{SRD} = \underbrace{E[Y_i(1) - Y_i(0) \mid X_i = c]}_{\text{Unobservable}} = \underbrace{\lim_{x \downarrow c} E[Y_i \mid X_i = x]}_{\text{Estimable}} - \underbrace{\lim_{x \uparrow c} E[Y_i \mid X_i = x]}_{\text{Estimable}}$$



Continuity/Extrapolation: Local Polynomial Methods

- Global polynomial regression: **not recommended**.
  - Runge's Phenomenon, counterintuitive weights, overfitting, lack of robustness.

- Local polynomial regression: captures idea of "localization".

Choose low poly order (p) and weighting scheme ($K(\cdot)$)

⇩

Choose bandwidth h: MSE-optimal or CE-optimal

⇩

Construct point estimator $\hat{\tau}$
(MSE-optimal $h \Longrightarrow$ optimal estimator)

⇩

Conduct robust bias-corrected inference
(CE-optimal $h \Longrightarrow$ optimal distributional approximation)

Local Polynomial Methods

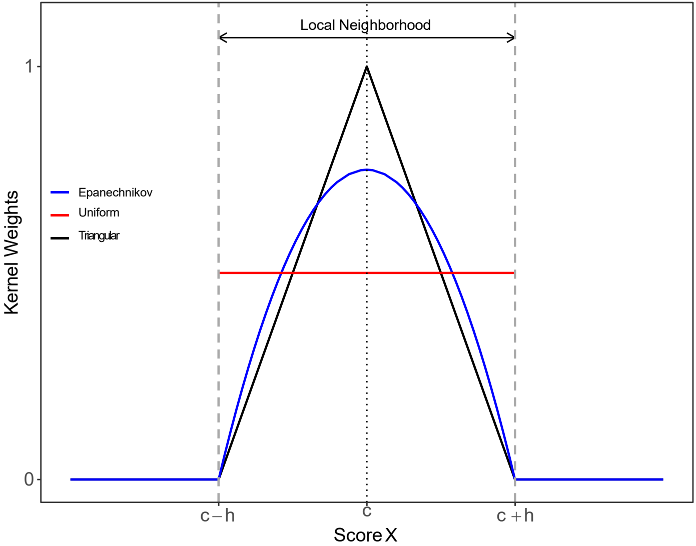
- **Idea**: approximate regression functions for control and treatment units *locally*.
- "Local-linear" ($p = 1$) estimator (w/ weights $K(\cdot)$):

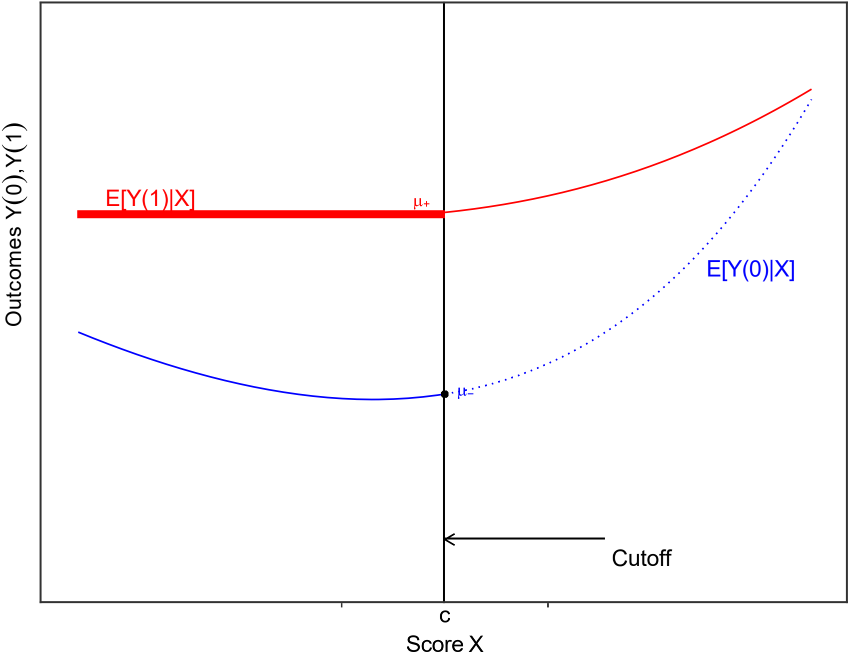
$-h \leq X_i < c$:	$c \leq X_i \leq h$:
$Y_i = \alpha_- + (X_i - c) \cdot \beta_- + \varepsilon_{-,i}$	$Y_i = \alpha_+ + (X_i - c) \cdot \beta_+ + \varepsilon_{+,i}$

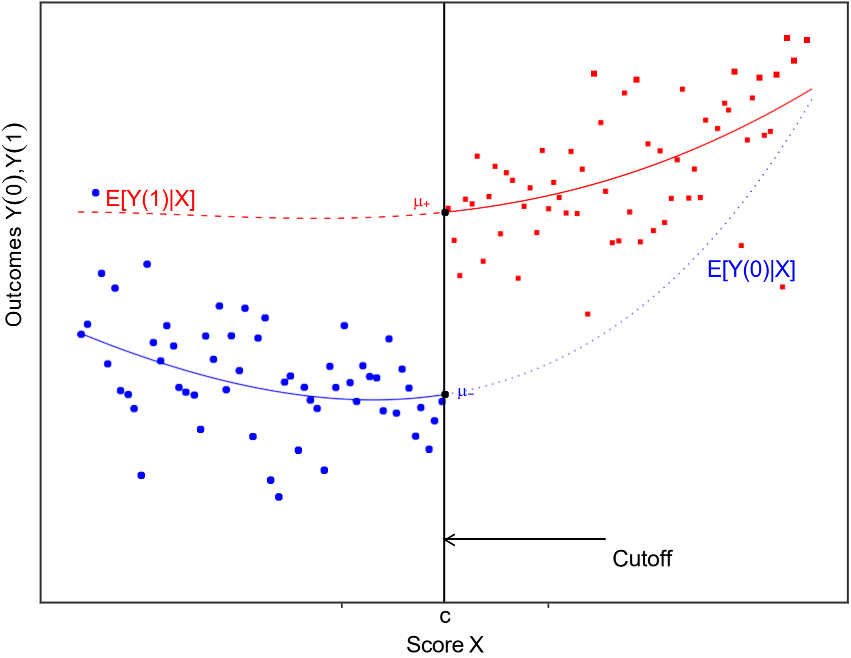
  - Treatment effect (at the cutoff): $\hat{\tau}_{SRD}(h) = \hat{\alpha}_+ - \hat{\alpha}_-$
- Can be estimated using linear models (w/ weights $K(\cdot)$):

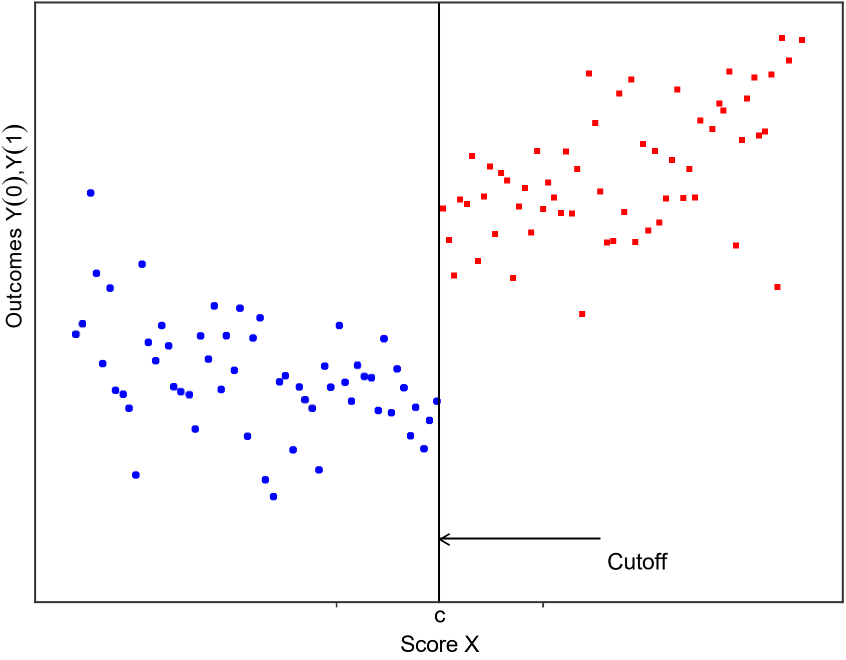
$$Y_i = \alpha + \tau_{SRD} \cdot T_i + (X_i - c) \cdot \beta_1 + T_i \cdot (X_i - c) \cdot \gamma_1 + \varepsilon_i, \qquad |X_i - c| \leq h$$

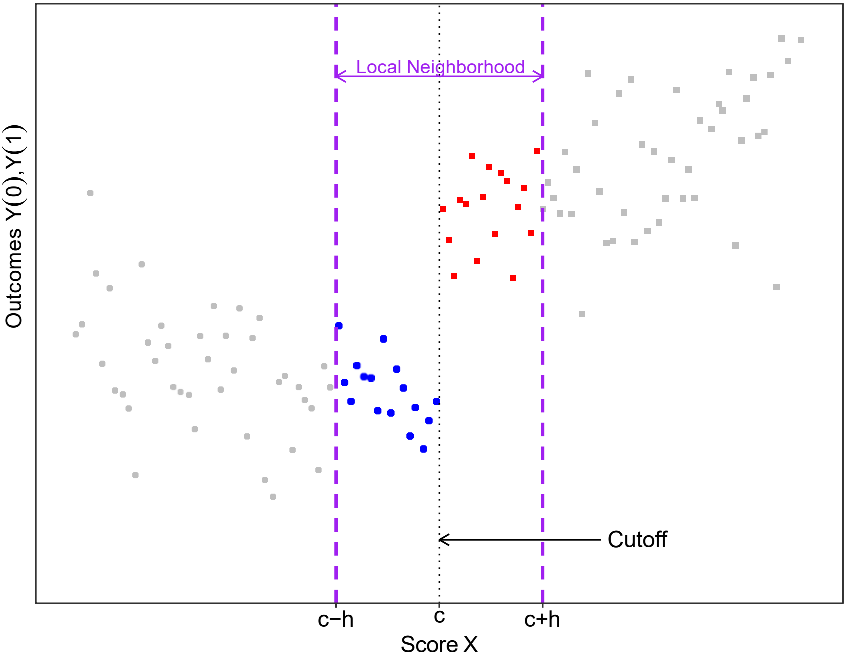
- Given p, K, h chosen $\Longrightarrow$ weighted least squares estimation.


Local Neighborhood
1
0
Kernel Weights
Epanechnikov
Uniform
Triangular
c − h
c
c + h
Score X

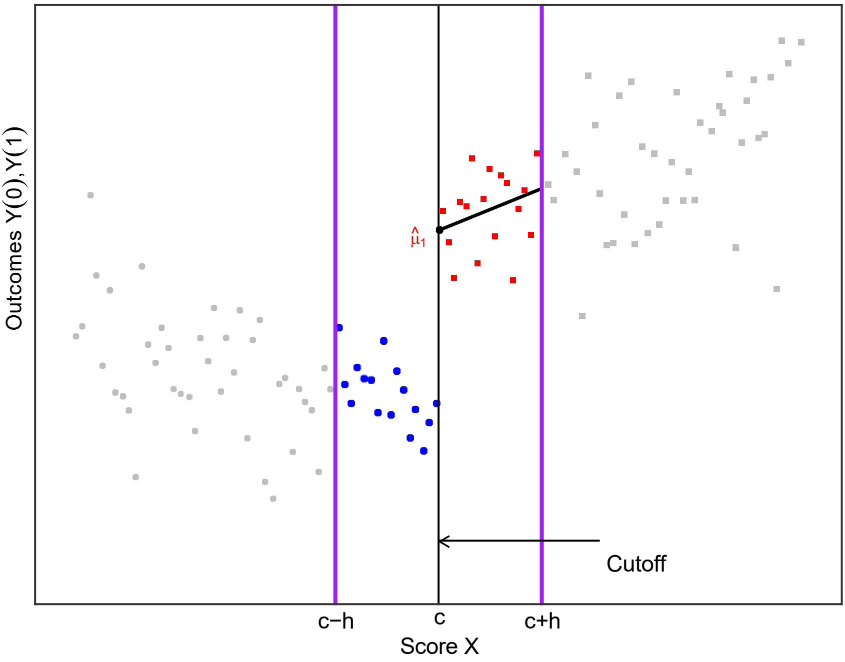

E[Y(1)|X]
μ+
E[Y(0)|X]
μ−
Cutoff
Outcomes Y(0),Y(1)
c
Score X


E[Y(1)|X]
μ_+
μ_-
E[Y(0)|X]
Outcomes Y(0),Y(1)
Cutoff
c
Score X

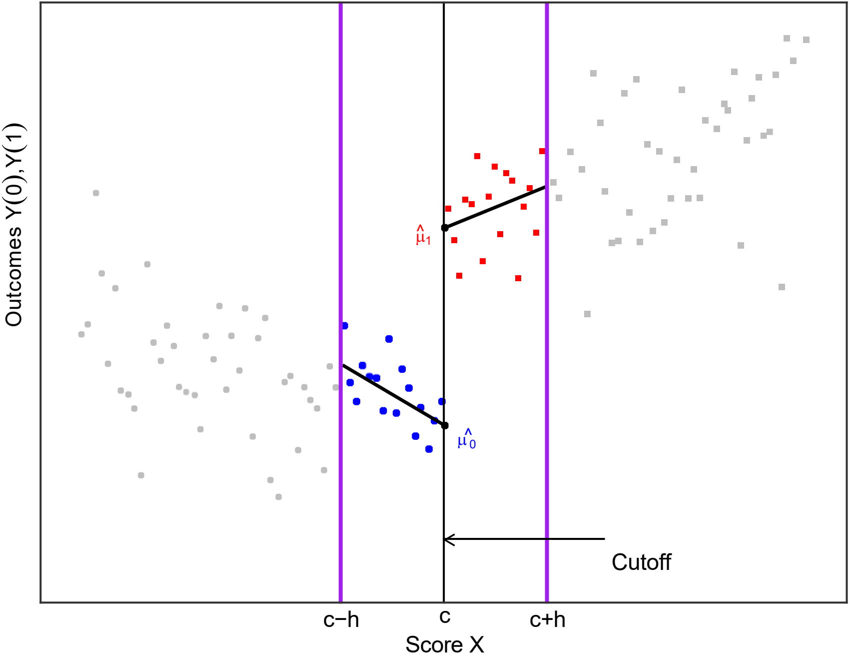

Outcomes Y(0),Y(1)
Cutoff
c
Score X



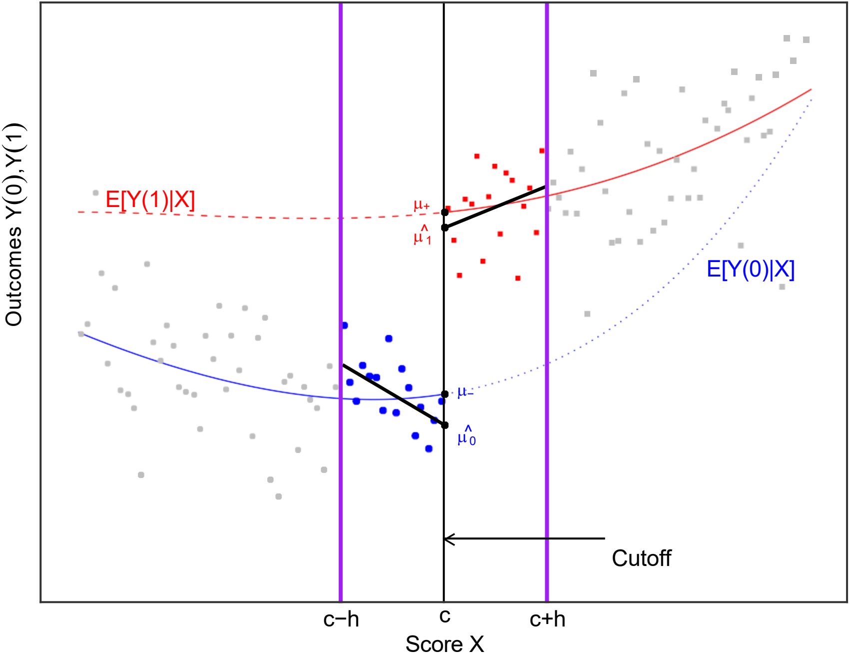
Local Neighborhood
Outcomes Y(0),Y(1)
Cutoff
c−h
c
c+h
Score X

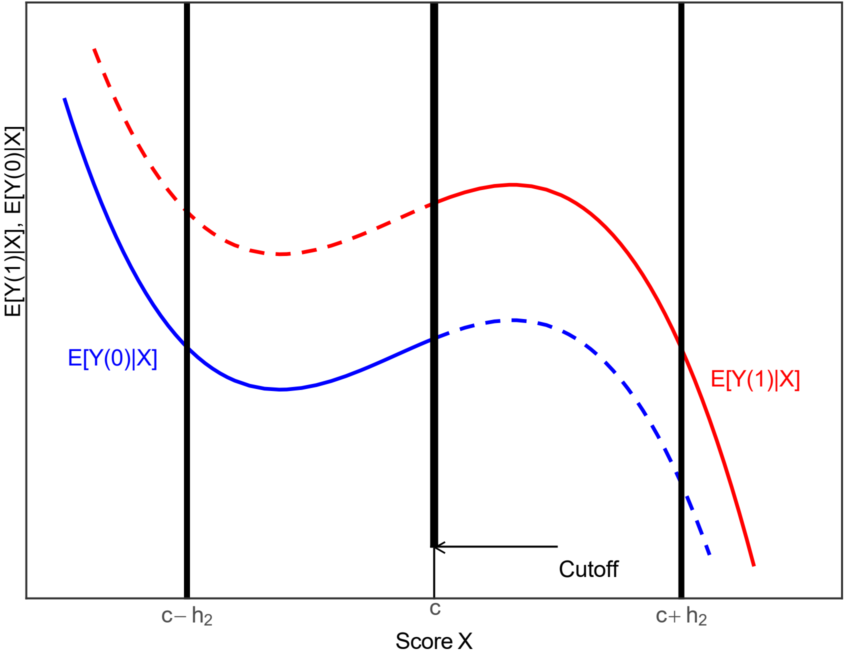


Outcomes Y(0),Y(1)
$\hat{\mu}_1$
Cutoff
c−h
c
c+h
Score X

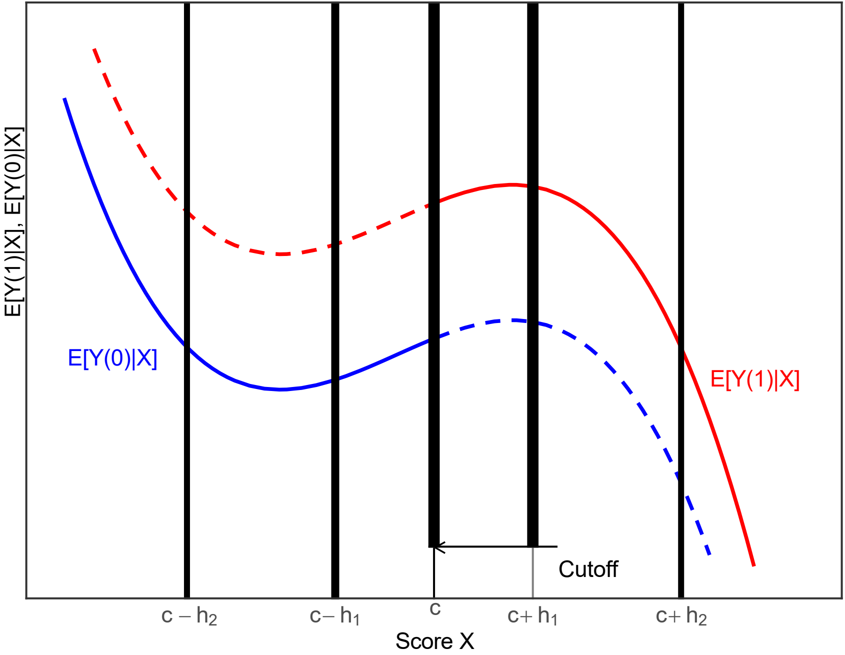


Outcomes Y(0),Y(1)
$\hat{\mu}_1$
$\hat{\mu}_0$
Cutoff
c−h
c
c+h
Score X


E[Y(1)|X]
E[Y(0)|X]
μ+
μ̂1
μ−
μ̂0
Cutoff
Outcomes Y(0),Y(1)
c−h
c
c+h
Score X



E[Y(1)|X], E[Y(0)|X]
E[Y(0)|X]
E[Y(1)|X]
Cutoff
c− h2
c
c+ h2
Score X


E[Y(1)|X], E[Y(0)|X]
E[Y(0)|X]
E[Y(1)|X]
Cutoff
c − h2
c− h1
c
c+ h1
c+ h2
Score X

Local Polynomial Methods: Choosing bandwidth ($p = 1$)

- Mean Square Error Optimal (MSE-optimal).

$$h_{MSE} = C_{MSE}^{1/5} \cdot n^{-1/5} \qquad C_{MSE} = C(K) \cdot \frac{\mathrm{Var}(\hat{\tau}_{SRD})}{\mathrm{Bias}(\hat{\tau}_{SRD})^2}$$

- Coverage Error Optimal (CE-optimal).

$$h_{CE} = C_{CE}^{1/4} \cdot n^{-1/4} \qquad C_{CE} = C(K) \cdot \frac{\mathrm{Var}(\hat{\tau}_{SRD})}{|\mathrm{Bias}(\hat{\tau}_{SRD})|}$$

- **Key idea**:
  - Trade-off bias and variance of $\hat{\tau}_{SRD}(h)$. Heuristically:

$$\uparrow \mathrm{Bias}(\hat{\tau}_{SRD}) \implies \downarrow \hat{h} \quad \text{and} \quad \uparrow \mathrm{Var}(\hat{\tau}_{SRD}) \implies \uparrow \hat{h}$$

  - Implementations: IK first-generation while CCT second-generation plug-in rule. They differ in the way $\mathrm{Var}(\hat{\tau}_{SRD})$ and $\mathrm{Bias}(\hat{\tau}_{SRD})$ are estimated.
  - Rule-of-thumb: $h_{CE} \propto n^{1/20} \cdot h_{MSE}$.

Conventional Inference Approach

- "Local-linear" ($p = 1$) estimator (w/ weights $K(\cdot)$):

$-h \leq X_i < c$:	$c \leq X_i \leq h$:
$Y_i = \alpha_- + (X_i - c) \cdot \beta_- + \varepsilon_{-,i}$	$Y_i = \alpha_+ + (X_i - c) \cdot \beta_+ + \varepsilon_{+,i}$

  - Treatment effect (at the cutoff): $\hat{\tau}_{SRD}(h) = \hat{\alpha}_+ - \hat{\alpha}_-$

- Construct usual t-test. For $H_0 : \tau_{SRD} = 0$,

$$T(h) = \frac{\hat{\tau}_{SRD}}{\sqrt{\hat{V}}} = \frac{\hat{\alpha}_+ - \hat{\alpha}_-}{\sqrt{\hat{V}_+ + \hat{V}_-}} \approx_d N(0, 1)$$

- Naïve 95% Confidence interval:

$$I(h) = \hat{\tau}_{SRD} \pm 1.96 \cdot \sqrt{\hat{V}}$$

Robust Bias Correction Approach

- **Key Problem**:

$$T(h_{\mathrm{MSE}}) = \frac{\hat{\tau}_{\mathrm{SRD}}}{\sqrt{\hat{V}}} \approx_d \mathsf{N}(\mathsf{B}, 1) \neq \mathsf{N}(0, 1)$$

  - B captures bias due to misspecification error.

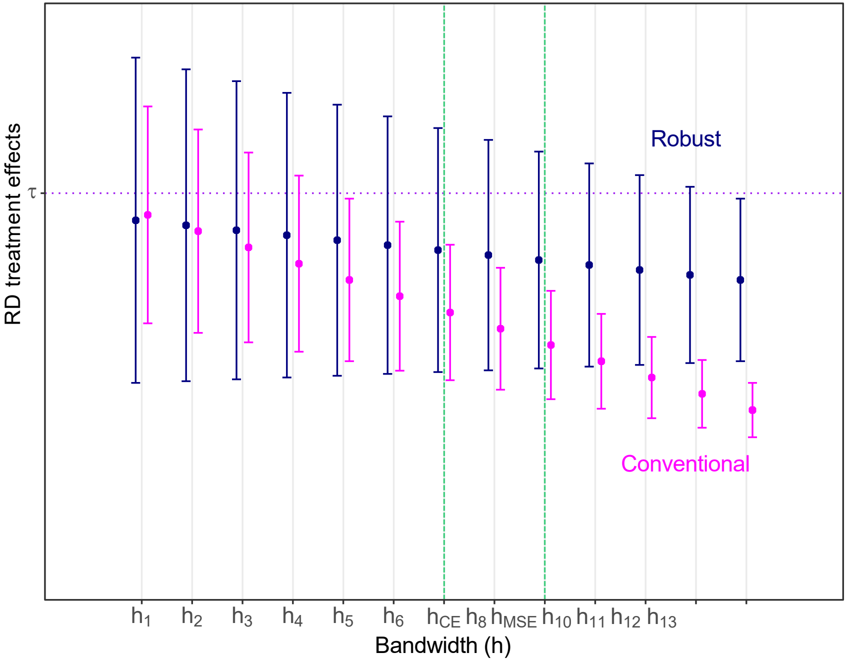
- **RBC distributional approximation**:

$$T^{\mathrm{bc}}(h) = \frac{\hat{\tau}_{\mathrm{SRD}} - \hat{\mathsf{B}}}{\sqrt{\hat{V}}} = \underbrace{\frac{\hat{\tau}_{\mathrm{SRD}} - \mathsf{B}}{\sqrt{\hat{V}}}}_{\approx_d \mathsf{N}(0,1)} + \underbrace{\frac{\mathsf{B} - \hat{\mathsf{B}}}{\sqrt{\hat{V}}}}_{\approx_d \mathsf{N}(0,\gamma)}$$

  - $\hat{\mathsf{B}}$ is constructed to estimate leading bias B, that is, misspecification error.

- **RBC** 95% **Confidence Interval**:

$$I_{\mathrm{RBC}} = \left(\hat{\tau}_{\mathrm{SRD}} - \hat{\mathsf{B}}\right) \pm 1.96 \cdot \sqrt{\hat{V} + \hat{W}}$$


RD treatment effects
τ
Robust
Conventional
h1 h2 h3 h4 h5 h6 hCE h8 hMSE h10 h11 h12 h13
Bandwidth (h)

Empirical Illustration: Head Start (Ludwig and Miller, 2007, QJE)

- **Problem**: impact of Head Start on Infant Mortality

- **Data**:

  Y_i = child mortality 5 to 9 years old

  T_i = whether county received Head Start assistance

  X_i = 1960 poverty index (c = 59.1984)

  Z_i = see database.

- **Potential outcomes**:

  $Y_i(0)$ = child mortality if **had not received** Head Start

  $Y_i(1)$ = child mortality if **had received** Head Start

- **Causal Inference**:

  $$Y_i(0) \neq Y_i | T_i = 0 \quad \text{and} \quad Y_i(1) \neq Y_i | T_i = 1$$

TABLE III
Regression Discontinuity Estimates of the Effect of Head Start Assistance on Mortality

Variable	Control mean	Nonparametric estimator			Parametric: Flexible linear	Parametric: Flexible quadratic
Bandwidth or poverty range		9	18	36	8	16
Number of observations (counties) with nonzero weight		527	961	2,177	484	863
Main results						
Ages 5–9, Head Start-related causes, 1973–1983	3.238	1.895	1.198	1.114	2.201	2.558
		(0.980)	(0.796)	(0.544)	(1.004)	(1.261)
		[0.036]	[0.081]	[0.027]	[0.022]	[0.021]
Specification checks						
Ages 5–9, injuries, 1973–1983	22.303	0.195	2.426	0.679	0.164	0.775
		(3.472)	(2.476)	(1.785)	(3.380)	(3.401)
		[0.924]	[0.345]	[0.755]	[0.998]	[0.835]
Ages 5–9, all causes, 1973–1983	40.232	3.416	0.053	1.537	3.896	2.927
		(4.311)	(3.098)	(2.253)	(4.268)	(4.295)
		[0.415]	[0.982]	[0.558]	[0.317]	[0.505]
Ages 25+, Head Start-related causes, 1973–1983	131.825	2.204	6.016	5.872	2.091	2.574
		(5.719)	(4.349)	(3.338)	(5.581)	(6.415)
		[0.700]	[0.147]	[0.114]	[0.749]	[0.689]

Outline

Local Randomization Approach to RD Design

- **Key assumption**: exists window $W = [c - w, c + w]$ around cutoff where subjects are as-if randomly assigned to either side of cutoff:

  1. Joint probability distribution of scores for units in the W is known:

     $$\mathsf{P}[\mathbf{X}_W \leq \mathbf{x}] = F(\mathbf{x}), \qquad \text{for some known joint c.d.f. } F(\mathbf{x}),$$

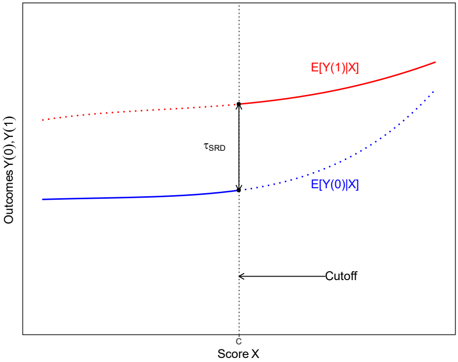
     where $\mathbf{X}_W$ denotes the vector of scores for all i such that $X_i \in W$.

  2. Potential outcomes not affected by value of the score:

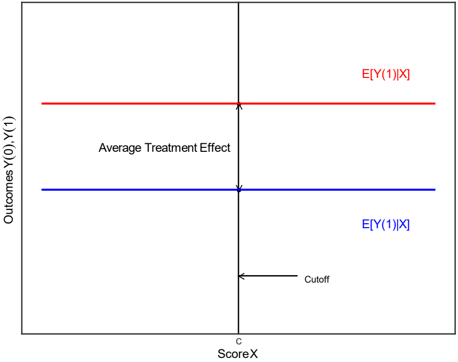
     $$Y_i(0, x) = Y_i(0),$$
     $$Y_i(1, x) = Y_i(1), \qquad \text{for all } X_i \in W.$$

- Note: stronger assumption than continuity-based approach.
  - Potential outcomes are a constant function of the score (can be relaxed).
  - Regression functions are not only continuous at c, but also completely unaffected by the running variable in W.

Experiment versus RD Design

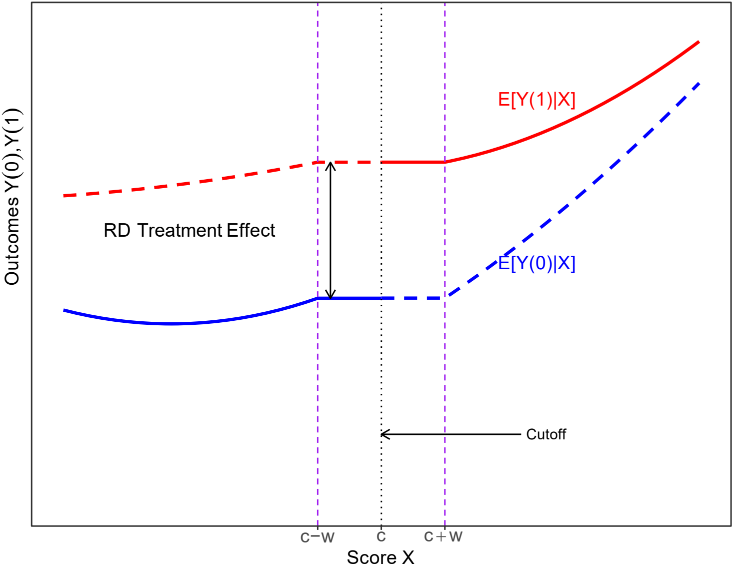


(a) RD Design



(b) Randomized Experiment

Local Randomization RD



Local Randomization Framework

- **Key idea**: exists window $W = [c - w, c + w]$ around cutoff where subjects are as-if randomly assigned to either side of cutoff.

- **Two Steps** (analogous to local polynomial methods):

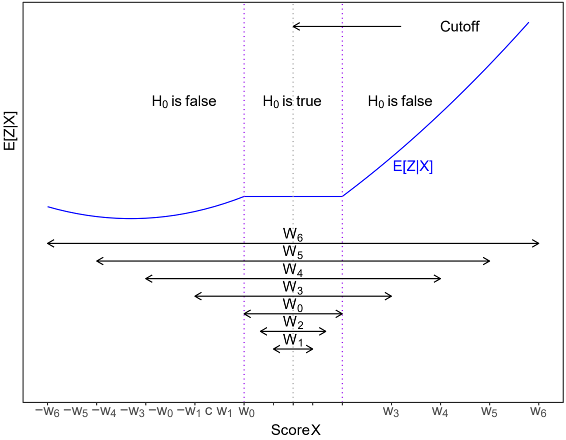
  1. Select window W.
  2. Given window W, perform estimation and inference.

- **Challenges**

  - Window (neighborhood) selection.
  - As-if random assumption good approximation *only very near cutoff*
  - Small sample.

Step 1: Choose the window W

- Find neighborhood where (pre-intervention) covariate-balance holds.
- Find neighborhood where outcome and score independent.
- Domain-specific or application-specific choice.



Step 2: Finite-sample and Large-sample Methods in W

- Given W where local randomization holds:
  - Randomization inference (Fisher): sharp null, finite-sample exact.
  - Design-based (Neyman): large-sample valid, conservative.
  - Large-sample standard: random potential outcomes, large-sample valid.

- All methods require window (W) selection, and choice of statistic.
  First two also require choice/assumptions assignment mechanism.
  Covariate-adjustments (score or otherwise) possible.

Empirical Illustration: Head Start (Ludwig and Miller, 2007, QJE)

- **Problem**: impact of Head Start on Infant Mortality

- **Data**:

  $Y_i =$ child mortality 5 to 9 years old

  $T_i =$ whether county received Head Start assistance

  $X_i =$ 1960 poverty index ($c = 59.1984$)

  $Z_i =$ see database.

- **Potential outcomes**:

  $Y_i(0) =$ child mortality if **had not received** Head Start

  $Y_i(1) =$ child mortality if **had received** Head Start

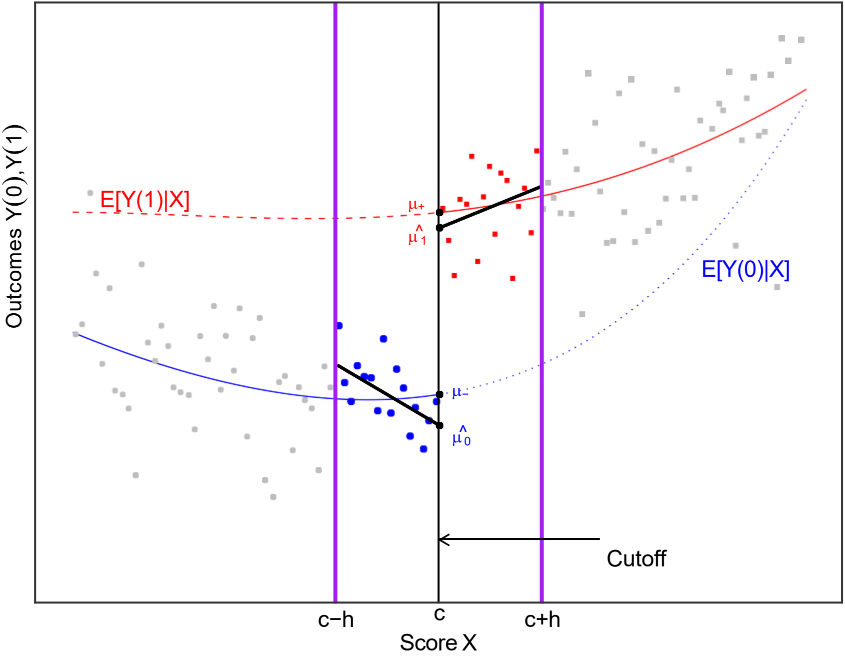
- **Causal Inference**:

  $Y_i(0) \neq Y_i | T_i = 0$ and $Y_i(1) \neq Y_i | T_i = 1$

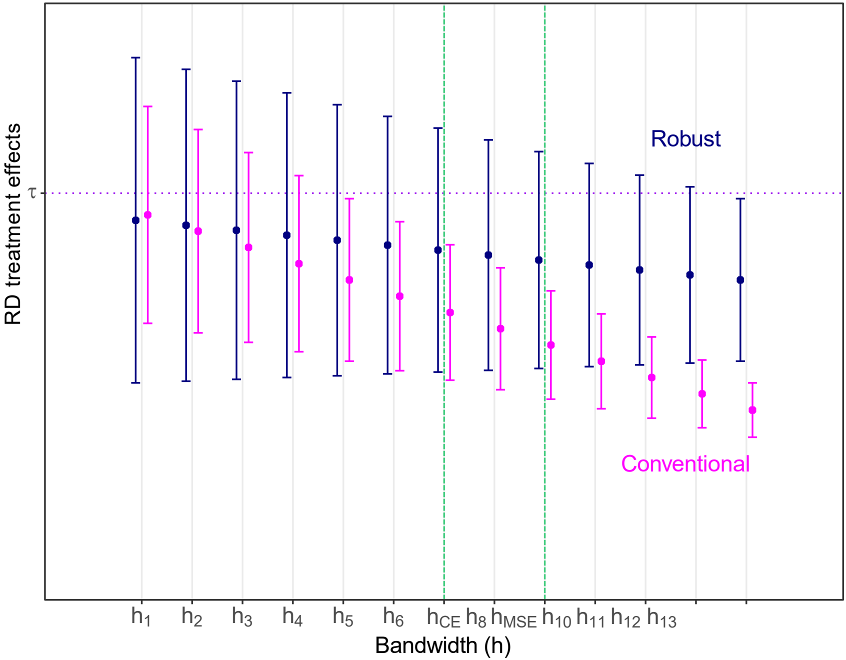
Outline

Falsification and Validation

- **RD plots and related graphical methods**:
  - Always plot data: main advantage of RD designs. (Check if RD design!)
  - Plot histogram of X_i (score) and its density. Careful: boundary bias.
  - RD plot $\mathrm{E}[Y_i|X_i = x]$ (outcome) and $\mathrm{E}[Z_i|X_i = x]$ (pre-intervention covariates).
  - Be careful not to oversmooth data/plots.
- **Sensitivity and related methods**:
  - Score density continuity: binomial test and continuity test.
  - Pre-intervention covariate no-effect (covariate balance).
  - Placebo outcomes no-effect.
  - Placebo cutoffs no-effect: informal continuity test away from c.
  - Donut hole: testing for outliers/leverage near c.
  - Different bandwidths: testing for misspecification error.
  - Many other setting-specific (fuzzy, geographic, etc.).



E[Y(1)|X]
E[Y(0)|X]
μ+
μ̂1
μ−
μ̂0
Cutoff
c−h
c
c+h
Score X
Outcomes Y(0),Y(1)


RD treatment effects
τ
Robust
Conventional
h1 h2 h3 h4 h5 h6 hCE h8 hMSE h10 h11 h12 h13
Bandwidth (h)

Empirical Illustration: Head Start (Ludwig and Miller, 2007, QJE)

- **Problem**: impact of Head Start on Infant Mortality

- **Data**:

  $Y_i =$ child mortality 5 to 9 years old

  $T_i =$ whether county received Head Start assistance

  $X_i =$ 1960 poverty index ($c = 59.1984$)

  $Z_i =$ see database.

- **Potential outcomes**:

  $Y_i(0) =$ child mortality if **had not received** Head Start

  $Y_i(1) =$ child mortality if **had received** Head Start

- **Causal Inference**:

  $Y_i(0) \neq Y_i | T_i = 0$ and $Y_i(1) \neq Y_i | T_i = 1$