Regression Discontinuity Design to Measure Incumbency Advantage in US Congressional Elections

STAT530 Final Project

Prayag & Tyler

4/21/23

Presentation Outline

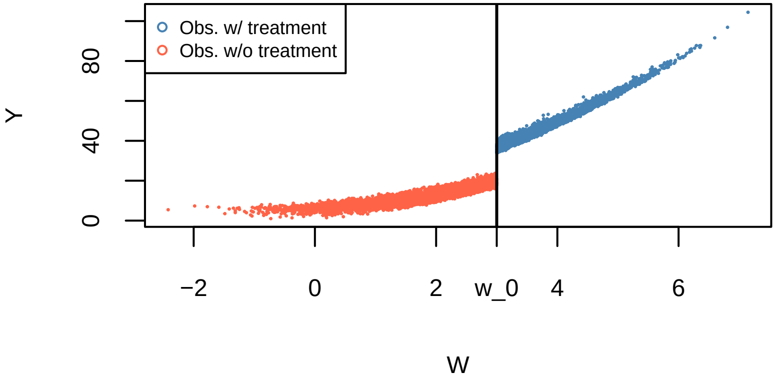
- Introduction to Regression Discontinuity Design
- [ADD LATER]
- [ADD LATER]
- [ADD LATER]

Regression Discontinuity Design

Introduction: regression discontinuity design (RDD) is a quasi-experimental pretest-posttest design that aims to determine the causal effects of interventions by assigning a cutoff or threshold above or below which an intervention is assigned.

Regression Discontinuity Design

Goal: Try to estimate the causal effect of intervention/treatment, $\tau(W = w_0)$, at $W = w_0$ on Y with covariate X.

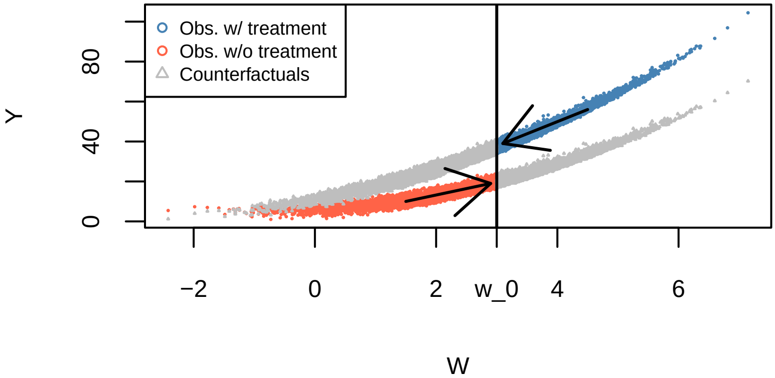


Regression Discontinuity Design

Continuity Assumption: Assume that the conditional expectation of potential outcome at the threshold $W = w_0$ can be approximated well from one side rather than from both sides, i.e.,

$$\mathbb{E}(Y_i^1 | W_i = w_0) = \lim_{w \downarrow w_0} \mathbb{E}(Y_i^1 | W_i = w),$$

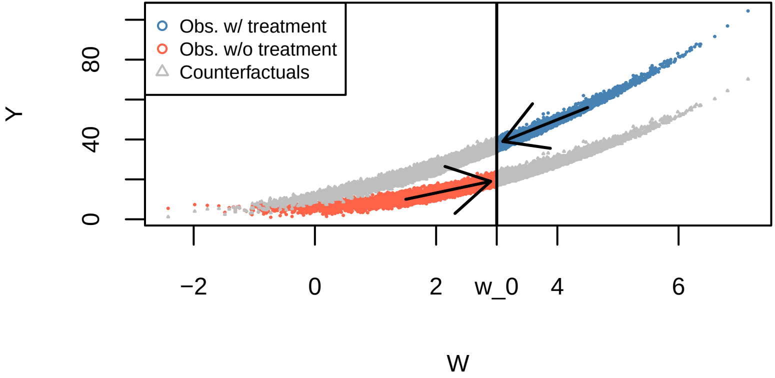
$$\mathbb{E}(Y_i^0 | W_i = w_0) = \lim_{w \uparrow w_0} \mathbb{E}(Y_i^0 | W_i = w).$$



Regression Discontinuity Design

Therefore, under the continuity of potential outcomes assumption, we can identify the ATE, $\tau(w_0)$, at the threshold $W = w_0$ as

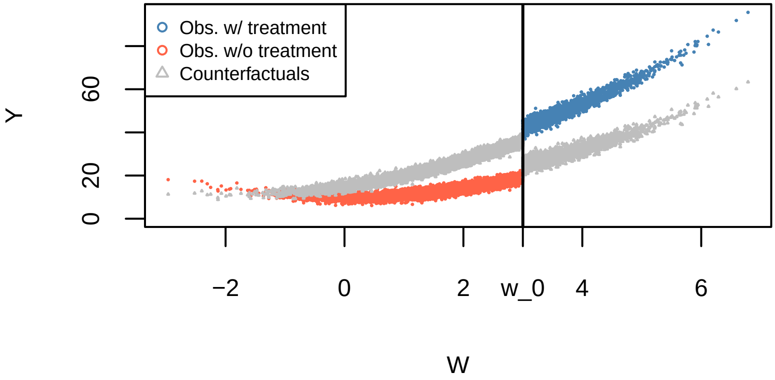
$$\begin{aligned} \text{ATE} &= \mathbb{E}(Y_i^1 - Y_i^0 | W_i = w_0) \\ &= \lim_{w \downarrow w_0} \mathbb{E}(Y_i^1 | W_i = w) - \lim_{w \uparrow w_0} \mathbb{E}(Y_i^0 | W_i = w) \end{aligned}$$



Regression Discontinuity Design

A **violation** of the continuity assumption: the barely-controls and the barely-treated come from very different populations, thus our RRD is unknowingly invalidated, since we will overestimate $\tau(w_0)$.

Thus, merely plotting observed Y versus W, we have no way of knowing if or if not the continuity assumption holds at the threshold $W = w_0$.



Regression Discontinuity Design

We can test for validity of the continuity assumption using our observed data:

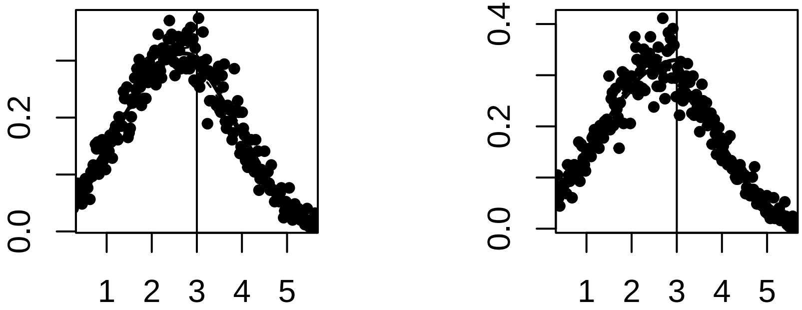
- Lack of discontinuity in desnity of assignment variable W at threshold $W = w_0$ (McCrary density test).
- Lack of discontinuity in pretreatment covariates conditional density $X|W$ at the threshold $W = w_0$ then represents empirical evidence for the continuity of expected potential outcomes so long as we test all relevant pretreatment covariates, i.e., $f(X = x|W = w)$ must be continuous in w at $w = w_0$ for all $x \in X$. (Regress X on W at threshold.)

Regression Discontinuity Design

- McCrary density test for density of assignment variable W for our two sets of data:

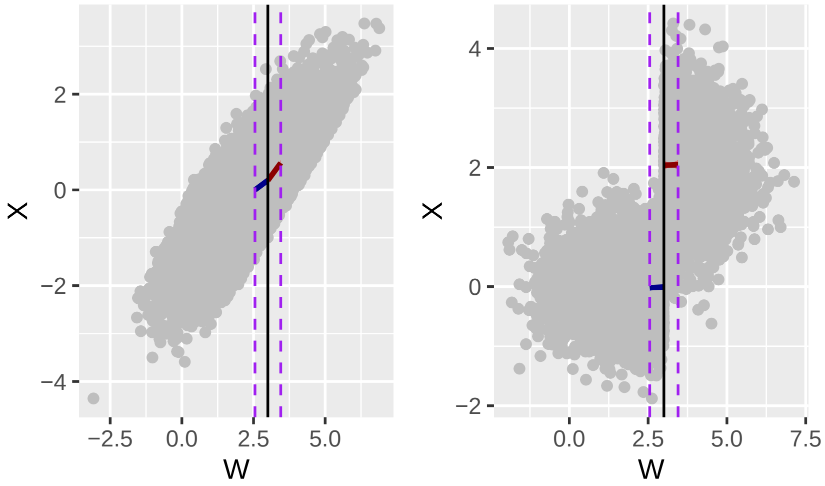
```
[1] 0.8468836

[1] 0.09318225
```



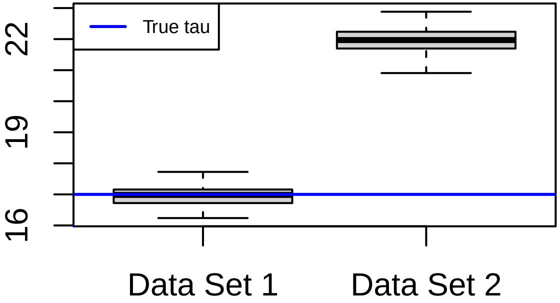
Regression Discontinuity Design

- Check for empirical discontinuity in covariate X's density conditional on the assignment variable W at the threshold using local linear regression for both data sets:



Regression Discontinuity Design

- Both data sets passed the McCrary test, data set 2 failed the test for conditional density continuity at the threshold, thus, blind RDD is not valid for data set 2, unless you incorporate the covariate in your model (but there are caveats to adding covariates, see Calonico et al., 2018).
- **Takeaway**: RDD will produce unbiased causal estimands of intervention at the threshold, $\tau(w_0)$, without taking into account the covariates but it requires our continuity assumption for ALL relevant covariates.



Regression Discontinuity Design

The estimator for the casual estimated at the assignment threshold $W = w_0$ is defined as

$$\hat{\tau}_{\mathsf{LLR}}(Y; W, K, b) = \hat{\alpha}_1 - \hat{\alpha}_0$$

$$(\hat{\alpha_0}, \hat{\beta_0}) = \underset{\alpha_0, \beta_0}{\arg\min} \sum_{i=1}^{n} 1\{w_l < W_i < w_0\}\{Y_i - \alpha_0 - \beta_0(W_i - w_0)\}^2 K(\cdot)$$

$$(\hat{\alpha_1}, \hat{\beta_1}) = \underset{\alpha_1, \beta_1}{\arg\min} \sum_{i=1}^{n} 1\{w_0 < W_i < w_u\}\{Y_i - \alpha_1 - \beta_1(W_i - w_0)\}^2 K(\cdot)$$

where $K(\cdot)$ is the weighting or kernel function and b is the choice of bandwidth that controls the weighting scheme in this local linear regression.