

# Regression Discontinuity Design to Measure Inc incumbency Advantage in US Congressional Elections

STAT530 Final Project

Prayag & Tyler

4/21/23

## Presentation Outline

- ▶ Introduction to Regression Discontinuity Design
- ▶ RDD Applied to US Congressional Elections: The Setup
- ▶ RDD Applied to US Congressional Elections: Results
- ▶ RDD Applied to US Congressional Elections: Conclusions

## Presentation Outline

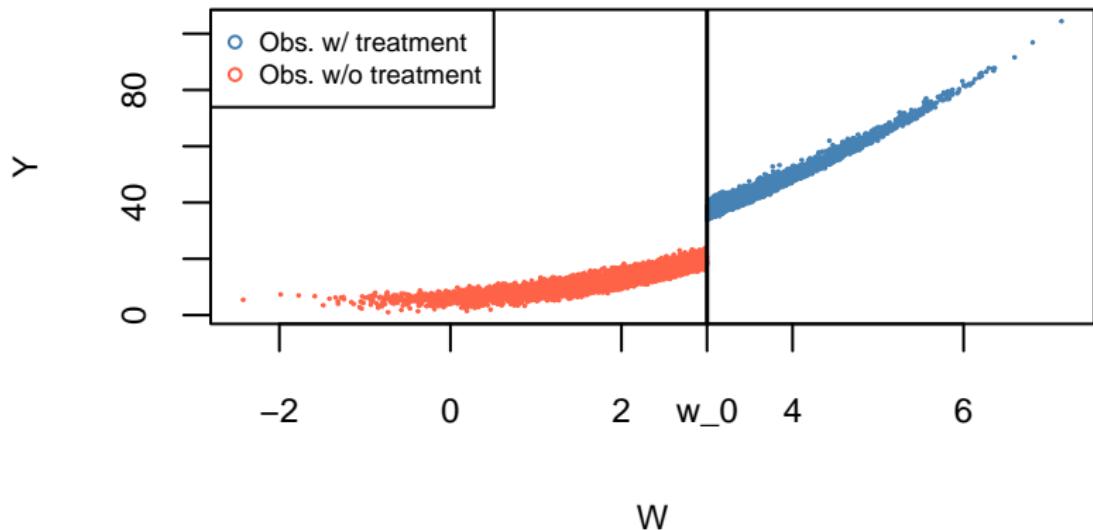
- ▶ Introduction to Regression Discontinuity Design
- ▶ RDD Applied to US Congressional Elections: The Setup
- ▶ RDD Applied to US Congressional Elections: Results
- ▶ RDD Applied to US Congressional Elections: Conclusions

# 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  $X$ .

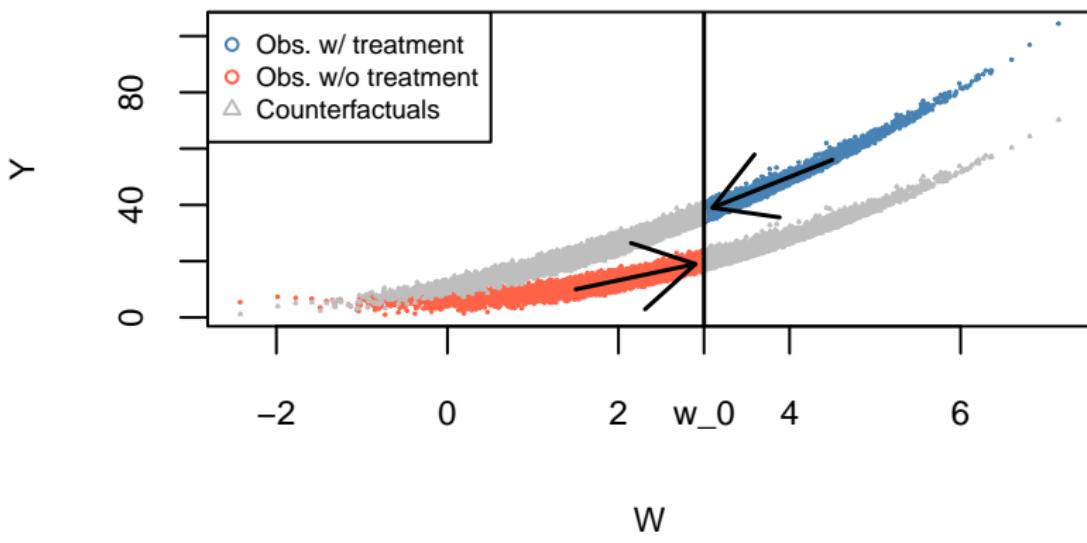


## Regression Discontinuity Design

**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.,

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

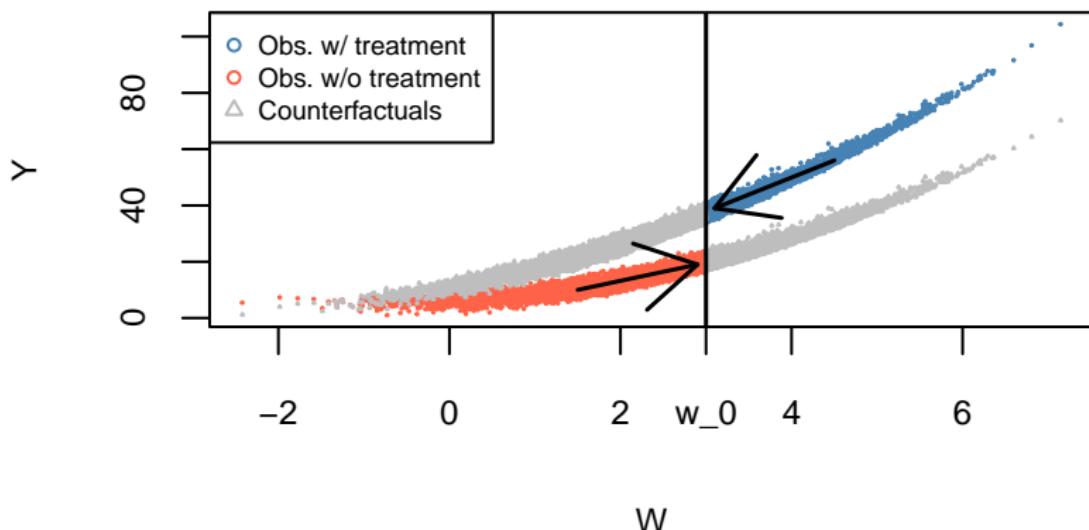
$$\mathbb{E}(Y_i^0|W_i = w_0) = \lim_{w \uparrow w_0} \mathbb{E}(Y_i^0|W_i = w).$$



## 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

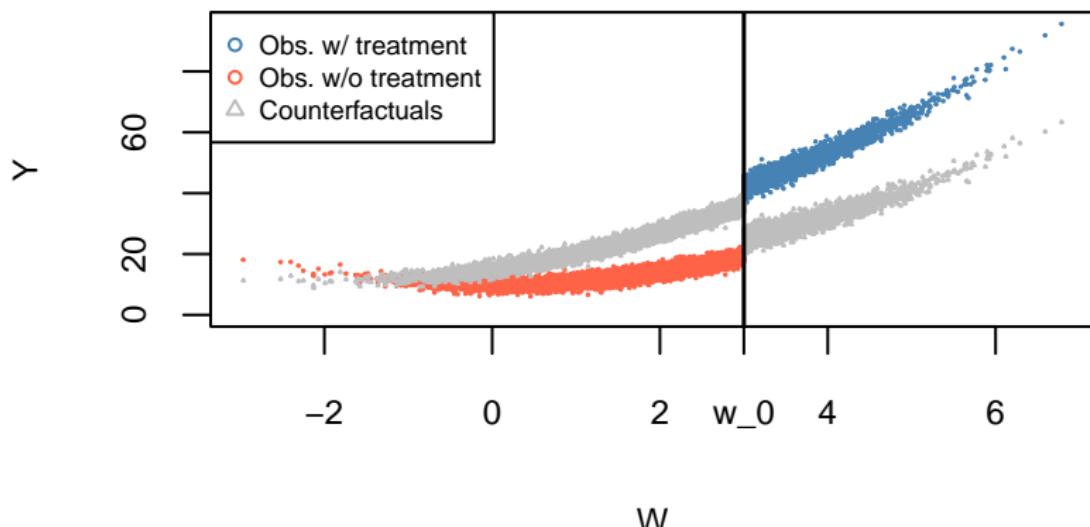
$$\begin{aligned}\text{ATE} &= \mathbb{E}(Y_i^1 - Y_i^0 | W_i = w_0) \\ &= \lim_{w \downarrow w_0} \mathbb{E}(Y_i^1 | W_i = w) - \lim_{w \uparrow w_0} \mathbb{E}(Y_i^0 | W_i = w)\end{aligned}$$



## 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:

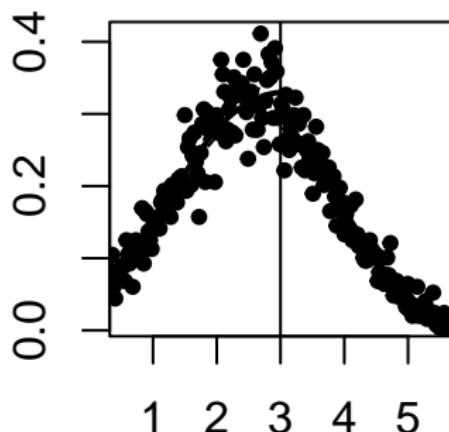
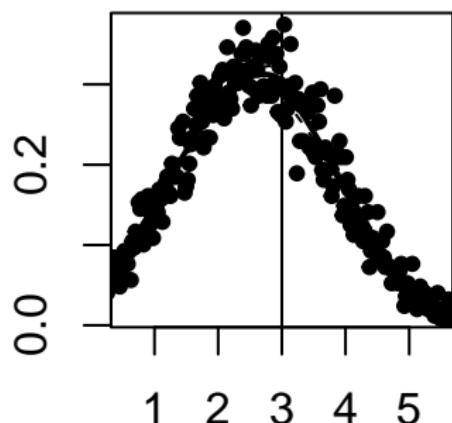
- ▶ Lack of discontinuity in density of assignment variable  $W$  at threshold  $W = w_0$  (McCrary density test).
- ▶ Lack of discontinuity in pretreatment covariates conditional density  $X|W$  at the threshold  $W = w_0$  then represents empirical evidence for the continuity of expected potential 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:

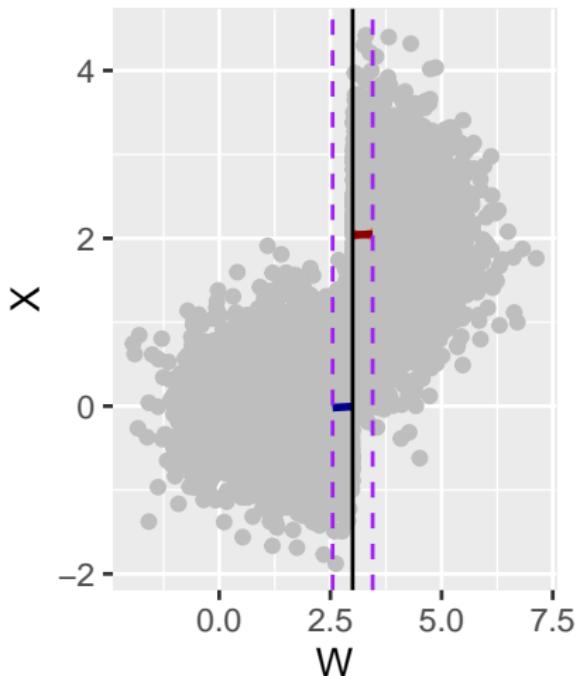
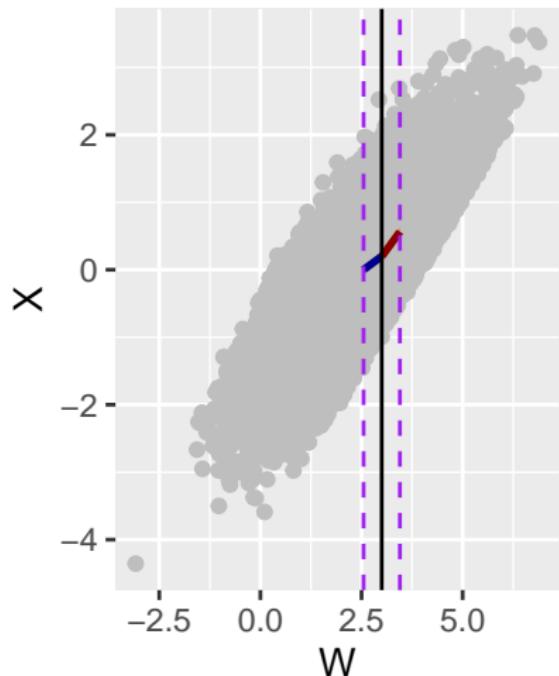
[1] 0.8468836

[1] 0.09318225



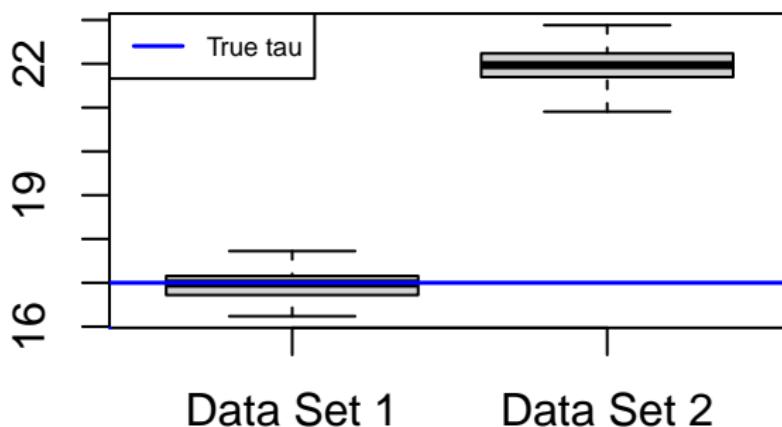
## Regression Discontinuity Design

- ▶ 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 causal effect at the assignment threshold  $W = w_0$  is defined as

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

$$(\hat{\alpha}_0, \hat{\beta}_0) = \arg \min_{\alpha_0, \beta_0} \sum_{i=1}^n 1\{w_l < W_i < w_0\} \{Y_i - \alpha_0 - \beta_0(W_i - w_0)\}^2 K(\cdot)$$

$$(\hat{\alpha}_1, \hat{\beta}_1) = \arg \min_{\alpha_1, \beta_1} \sum_{i=1}^n 1\{w_0 < W_i < w_u\} \{Y_i - \alpha_1 - \beta_1(W_i - w_0)\}^2 K(\cdot)$$

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.

## Presentation Outline

- ▶ Introduction to Regression Discontinuity Design
- ▶ **RDD Applied to US Congressional Elections: The Setup**
- ▶ RDD Applied to US Congressional Elections: Results
- ▶ RDD Applied to US Congressional Elections: Conclusions

## RRD Applied to US Congressional Elections: The Setup

### Project Motivations:

- ▶ One of the most striking facts of congressional politics in the United States is the consistently high rate of electoral success of incumbents, and the electoral advantage of incumbency is one of the most studied aspects of research on elections to the U.S. House (Gelman and King, 1990).
- ▶ The overwhelming success of House incumbents draws public attention whenever concerns arise that Representatives are using the privileges and resources of office to gain an “unfair” advantage over potential challengers.
- ▶ Def'n: The **incumbent** is the current holder of an office or position, usually in relation to an election.

# RRD Applied to US Congressional Elections: The Setup

Data Sources: Harvard Dataverse and MIT Election Data + Science Lab

- ▶ U.S. House 1976–2020 Elections (Data set 1)
  - ▶ MIT Election Data and Science Lab, 2017, “U.S. House 1976–2020”
- ▶ Historical Congressional Legislation and District Demographics 1972-2014 (Data set 2)
  - ▶ Foster-Molina, Ella, 2017, “Historical Congressional Legislation and District Demographics 1972-2014”



Dataverse

# RRD Applied to US Congressional Elections: The Setup

Observed variables and covariates after data wrangling:

Data Set 1	Data Set 2
Year	Year
State	State
District	District
<b>Dem. Vote Margin</b>	Unemployment
Candidate	Race
	Education

**Dem. Vote Margin** = (Dem. Vote Share) - (Greatest Non-Dem. Vote Share)

# RRD Applied to US Congressional Elections: The Setup

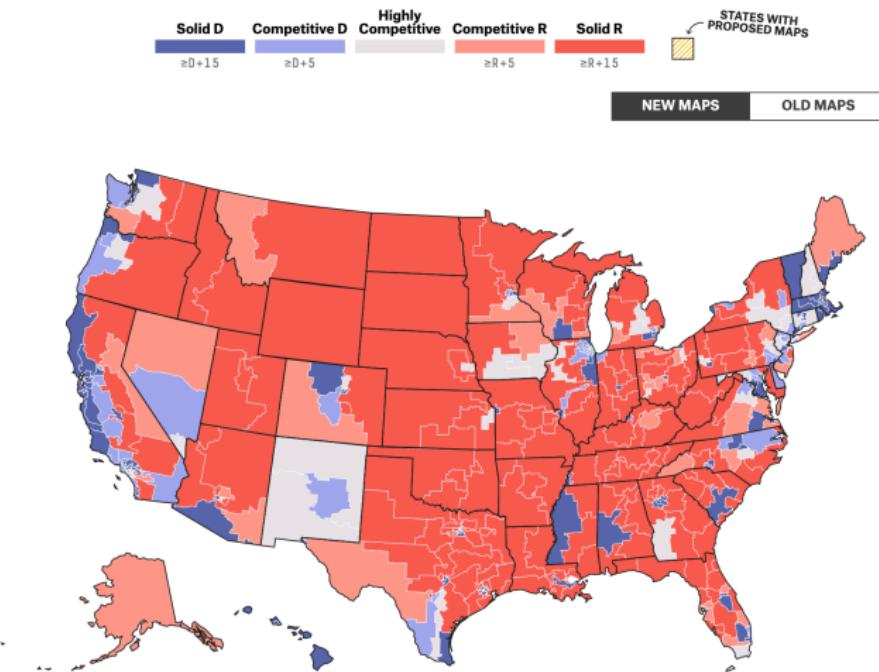


Figure 1: 2022 Congressional Districts and Partisan Lean

## Presentation Outline

- ▶ Introduction to Regression Discontinuity Design
- ▶ RDD Applied to US Congressional Elections: The Setup
- ▶ **RDD Applied to US Congressional Elections: Results**
- ▶ RDD Applied to US Congressional Elections: Conclusions