# Regression Discontinuity Designs and Difference-in-Differences

Jake Bowers, Ben Hansen & Tom Leavitt

August 12, 2018

## Contents

## 1 Regression Discontinuity Design

### 1.1 General Setup

Today we are going to use the data on close US House of Representatives races 1942–2008 used in **?**.[1] **?** engage in a debate whose participants seek to identify the causal effect of the so-called "incumbency advantage." That is, what effect does a candidate's status as an incumbent have on whether or not that candidate wins an election? Obviously, whether or not a candidate is an incumbent is *not* randomly assigned.

Let's first load the data:

```
rm(list = ls())

rdd_data <- read_dta("http://jakebowers.org/Matching/RDReplication.dta") %>% filter(Use == 1)  ## Use is indicator for
```

#### The Running Variable

The "running variable" is called DifDPct, which is defined as the Democratic margin of victory or defeat in the election; in other words, DifDPct is the difference between the percentage of all votes that were cast for the leading Democrat in the race and the percentage cast for the leading non-Democrat. Races in which no Democrat ran or in which the top two vote-getters were both Democrats are coded as missing.

[1] The full replication data is available for download here. But we read it directly below.

```
running_var <- matrix(c("DifDPct", "Democrat Margin of Victory"), ncol = 2, byrow = TRUE)

dimnames(running_var) <- list(1, c("Running Variable", "Description"))

kable(running_var)
```

| Running Variable | Description |
|---|---|
| DifDPct | Democrat Margin of Victory |

### The Treatment Variable

The treatment variable is whether or not the Democratic candidate wins the election or not. If the candidate wins the election, then that candidate is assigned to "treatment." If the candidate loses the election, then he or she is assigned to "control."

```
treatment <- matrix(c("DemWin", "Democrat Wins Election"), ncol = 2, byrow = TRUE)

dimnames(treatment) <- list(1, c("Treatment", "Description"))

kable(treatment)
```

| Treatment | Description |
|---|---|
| DemWin | Democrat Wins Election |

Now let's quickly look at the empirical distribution of the treatment variable:

```
table(rdd_data$DemWin)
```

| | Democrat Wins Election |
|---|---|
| 0 | 4507 |
| 1 | 5677 |

**Table 1:** The Treatment Variable

### Outcome Variables

In **?**, the primary outcome variables of interest are as follows: whether a democrat wins the next election, the proportion voting for a democrat in the next election, and the democratic vote margin in the next election.

```
dvs <- matrix(c("DWinNxt", "Dem Win t + 1", "DPctNxt", "Dem % t + 1", "DifDPNxt", "Dem % Margin t + 1"),
    ncol = 2, byrow = TRUE)

dimnames(dvs) <- list(seq(from = 1, to = 3, by = 1), c("Outcome", "Description"))

kable(dvs)
```

| Outcome | Description |
|---|---|
| DWinNxt | Dem Win t + 1 |
| DPctNxt | Dem % t + 1 |
| DifDPNxt | Dem % Margin t + 1 |

**Baseline Covariates**

The relevant baseline (i.e., pre-treatment) covariates about the races are:

| Covariate | Description |
|---|---|
| DWinPrv | Dem Win t - 1 |
| DPctPrv | Dem % t - 1 |
| DifDPPrv | Dem % Margin t - 1 |
| IncDWNOM1 | Inc's D1 NOMINATE |
| DemInc | Dem Inc in Race |
| NonDInc | Rep Inc in Race |
| PrvTrmsD | Dem's # Prev Terms |
| PrvTrmsO | Rep's # Prev Terms |
| RExpAdv | Rep Experience Adv |
| DExpAdv | Dem Experience Adv |
| ElcSwing | Partisan Swing |
| CQRating3 | CQ Rating {-1, 0, 1} |
| DSpndPct | Dem Spending % |
| DDonaPct | Dem Donation % |
| SoSDem | Dem Sec of State |
| GovDem | Dem Governor |
| DifPVDec | Dem Pres % Margin |
| DemOpen | Dem-held Open Seat |
| NonDOpen | Rep-held Open Seat |
| OpenSeat | Open Seat |
| VtTotPct | Voter Turnout % |
| GovWkPct | Pct Gov't Worker |
| UrbanPct | Pct Urban |
| BlackPct | Pct Black |
| ForgnPct | Pct Foreign Born |

## 1.2 Core RDD Assumptions

Let the index $i \in \{1, \ldots, n\}$ run over the $n$ experimental units. In the context of a regression discontinuity design, let $R_i$ denote the random "score variable" (also known as "running variable"). The random assignment variable, $Z_i \in \{0, 1\}$, which indicates whether subject $i$ is assigned to treatment, $Z_i = 1$, or to control, $Z_i = 0$, is a deterministic function of $R_i$. In the context of the incumbency advantage study, we can define $Z_i$ as follows:

$$Z_i \equiv \mathbf{I}\left[R_i > 0\right],$$