Regression Discontinuity Lab

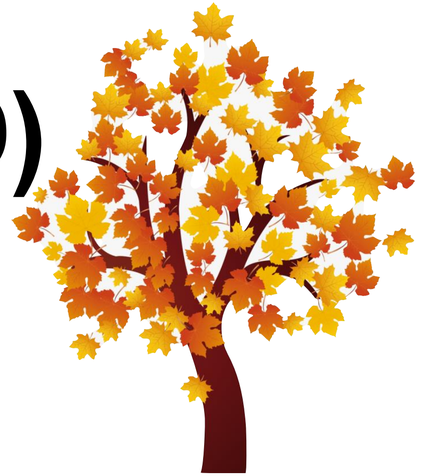
INFO / STSCI / ILRST 3900: Causal Inference

Wednesday, Oct 29

Reminders & Announcements

- Peer reviews for PS4 are due Tuesday, Nov 4
- Quiz 4 Tuesday, Nov 4
- Office hours:
    - Filippo: Thursday 4-5 pm in 321A CIS Building
    - Shira: Monday 5-6 pm in 329A CIS Building
    - Sam: Tuesday 4-5 pm, in 350 CIS Building

Regression Discontinuity Design (RDD)



Big Picture

- We use RDD when treatment is assigned according to some cutoff
  - Under the cutoff, you get no treatment
  - Above the cutoff, you get treatment
- Around the cutoff, we expect people to be similar
- Smooth potential outcomes at cutoff assumption: nothing but treatment should change at the cutoff
- Estimate a local average treatment effect (LATE): the effect of treatment on individuals **near the cutoff**

RDD Big Picture

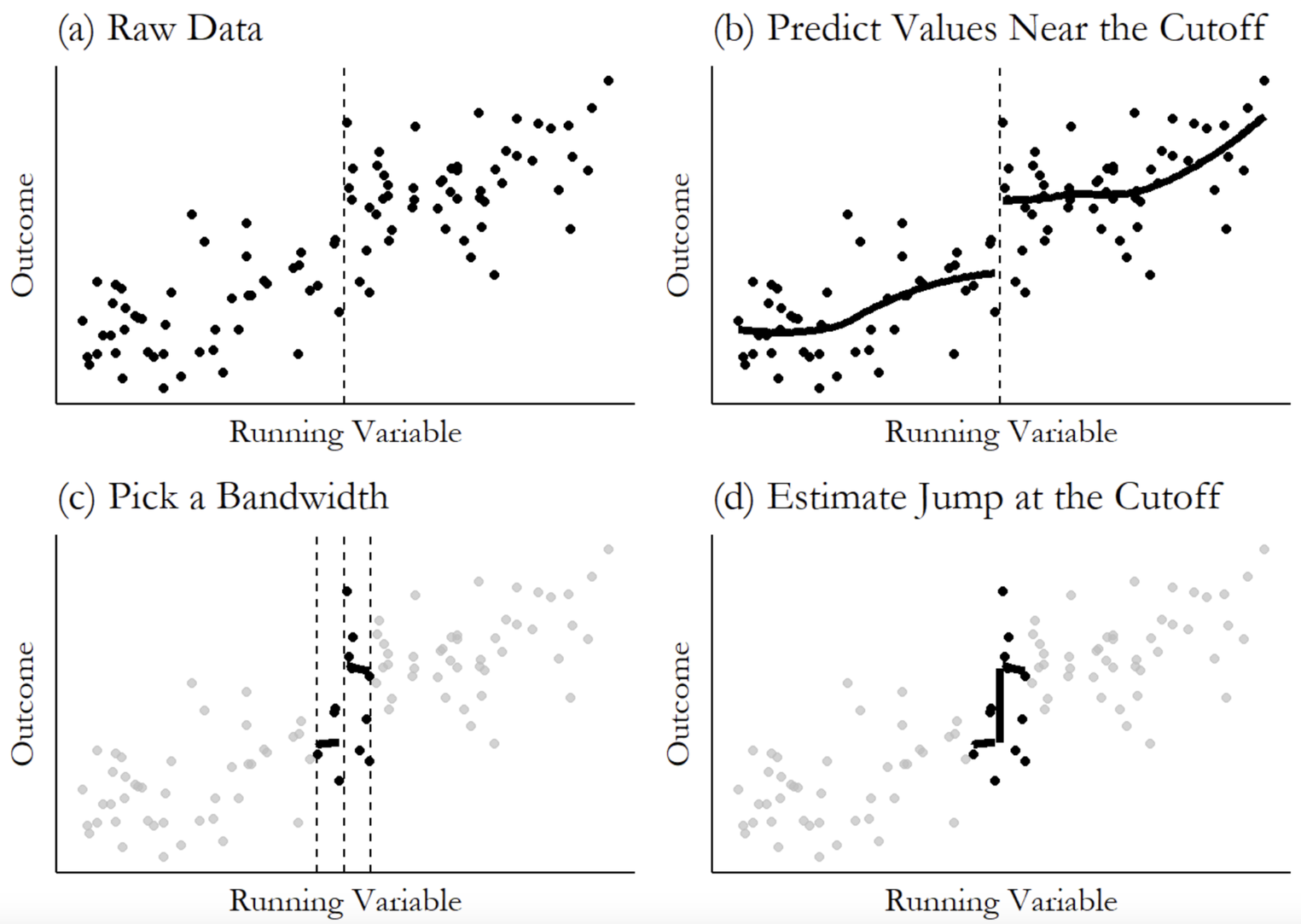
The Steps

(a) Collected data

(b) Predict values near the cutoff using regression models

(c) Determine how far away from the cutoff you're willing to look

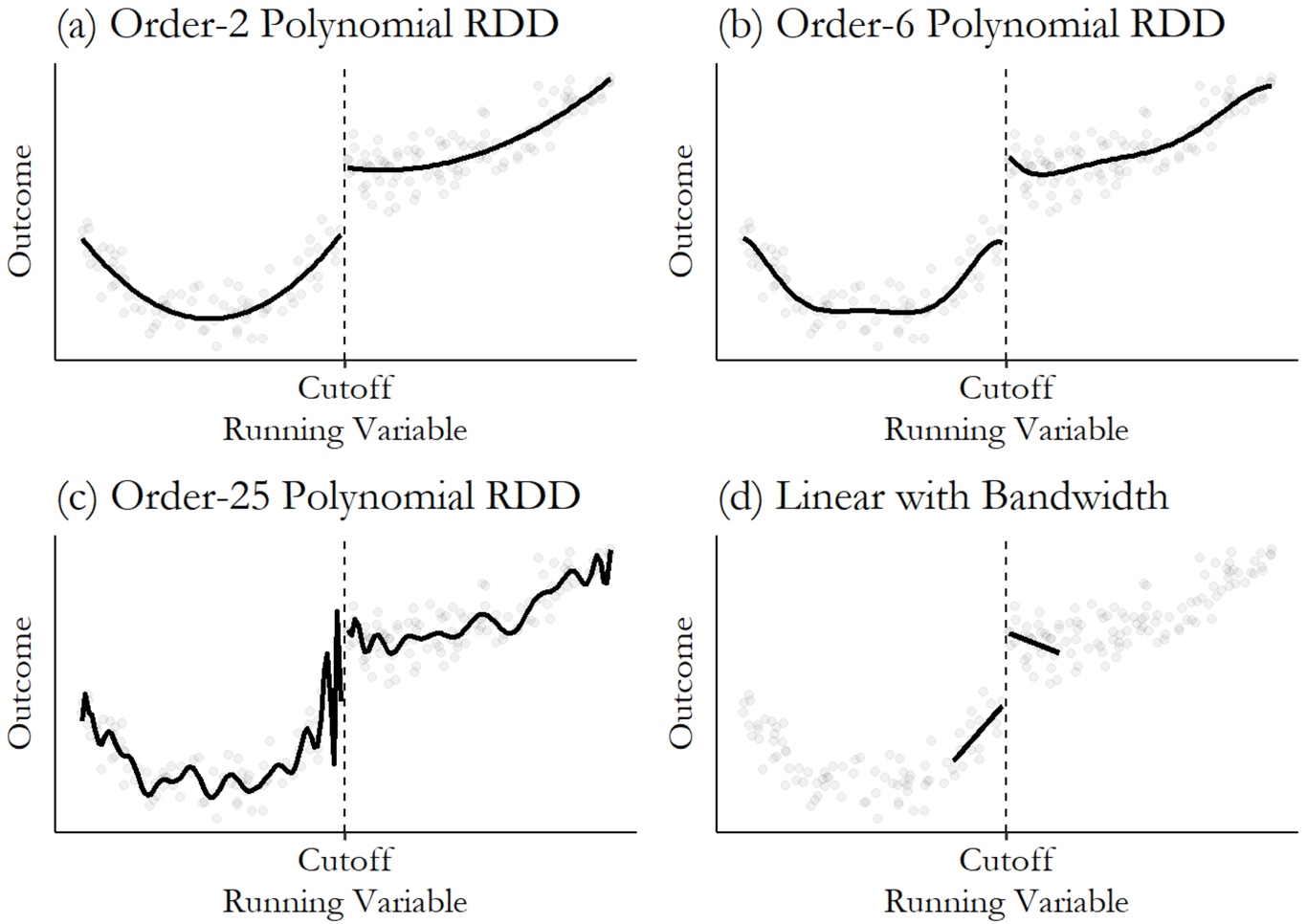
(d) Measure how far the jump is at the cutoff



Huntington-Klein, Nick. *The effect: An introduction to research design and causality*. Chapman and Hall/CRC, 2021.

Choosing a Bandwidth

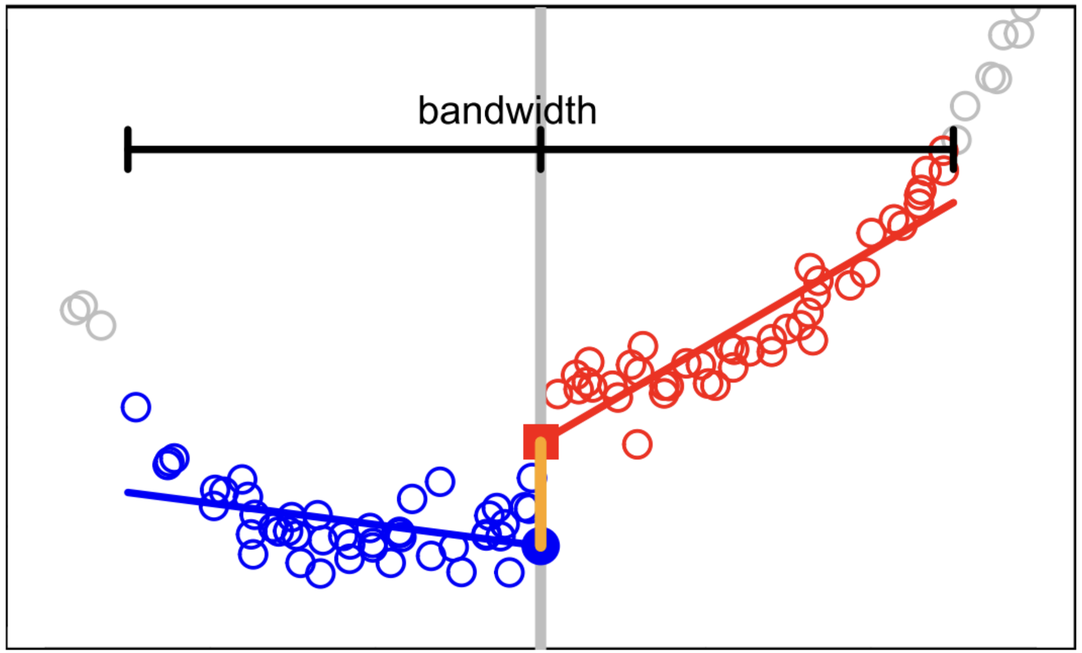
- Using linear regression, we're making the assumption $E[Y^{a=1}|X]$ is linear.
- What if it's not?
- Higher-degree models are prone to overfitting and high variance
- In practice, typically stick with linear but choose a bandwidth to trade off between bias and variance



Huntington-Klein, Nick. *The effect: An introduction to research design and causality*. Chapman and Hall/CRC, 2021.

Choosing a Bandwidth

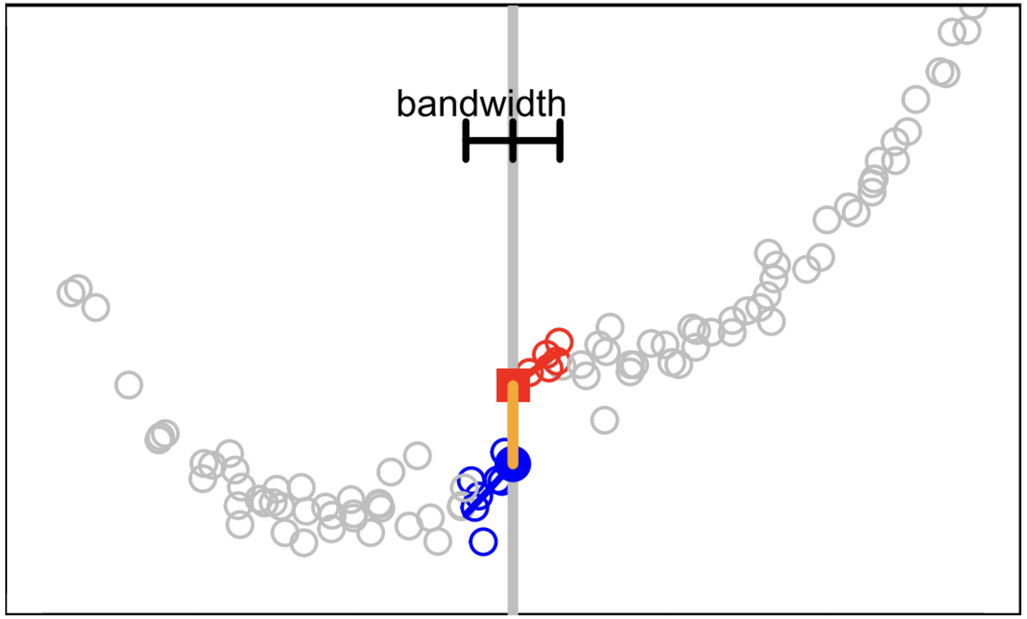
How do we choose a bandwidth?



- **Bias:** How far away are we from the true model?
- **Variance:** How much would my estimate change in a new sample?
- Larger bandwidth
  - **Increased bias** if we have an incorrect model, because results rely more on the function we assume (in this case “linear”)
  - **Decreased variance** because we use more more data since we include more individuals

Choosing a Bandwidth

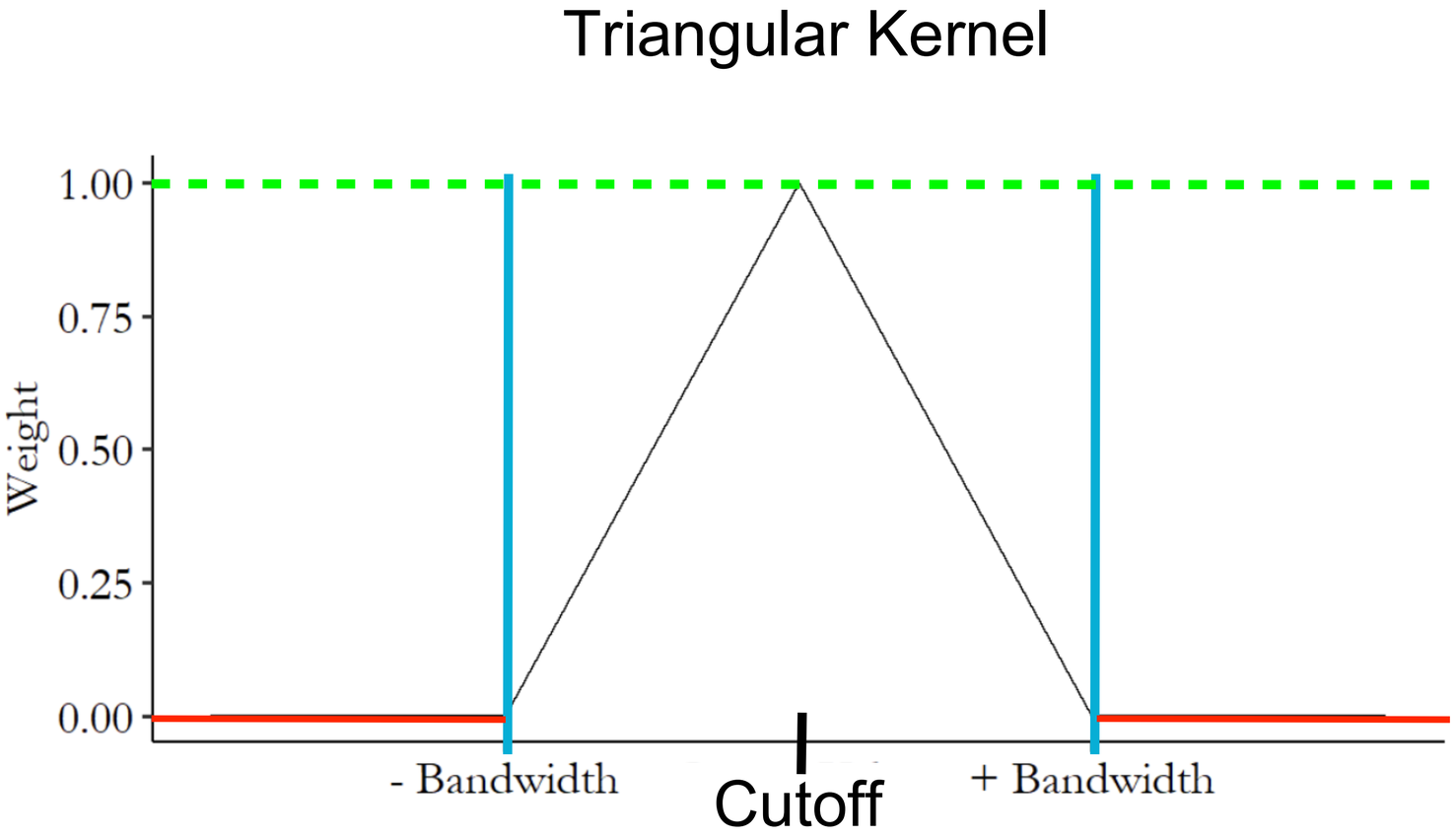
How do we choose a bandwidth?



- **Bias:** How far away are we from the true model?
- **Variance:** How much would my estimate change in a new sample?
- Smaller bandwidth
  - **Decreased bias**: We don't rely as much on the functional form we assumed
  - **Increased variance**: We have less data to work with
- Typically, bandwidth should decrease as sample size increases

Weighting

Big Picture



- Observations closest to the cutoff are the “most accurate”
- We can weight closer observations more heavily than farther observations
- Triangular kernel: different weights within the cutoff, depending on how close/far

RDD in Code

- The rdrobust package in R basically takes care of everything for us!
  - Chooses bandwidth, estimates causal effects, gives standard error
  - results <- rdrobust(y, x, kernel, p, h)
    - y = outcome, x = running variable, kernel = weighting (optional), p = degree of polynomial (default: linear), h = bandwidth (optional)

```
# uniform kernel with bandwidth 10
out <- rdrobust(dem_vote_t2, dem_margin_t0, kernel = 'uniform',  p = 1, h = 10)
summary(out)
```

These are parameters you can play around with to explore a bias-variance trade-off

Regression Discontinuity in Code

The rdrobust package in R

```
# uniform kernel with bandwidth 10
out <- rdrobust(dem_vote_t2, dem_margin_t0, kernel = 'uniform',  p = 1, h = 10)
summary(out)
```

```
## Number of Obs.                 1297
## BW type                      Manual
## Kernel                      Uniform
## VCE method                       NN
##
## Number of Obs.                  595          702
## Eff. Number of Obs.             245          206
## Order est. (p)                    1            1
## Order bias  (q)                   2            2
## BW est. (h)                  10.000       10.000
## BW bias (b)                  10.000       10.000
## rho (h/b)                     1.000        1.000
## Unique Obs.                     595          702
##
## ============================================================================
##         Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## ============================================================================
##   Conventional     6.899     1.722     4.007     0.000     [3.525 , 10.273]
##         Robust         -         -     3.891     0.000     [5.156 , 15.624]
## ============================================================================
```

Causal effect estimate!

RMarkdown Activity

An Application to Party Advantages in the U.S. Senate

- Your activity for today is to go through the notebook and answer some questions
- You can work with people around you
- Try to finish the first 3 sections, and move on to section 4 if you have extra time
- Goals:
    - Interact with RDD through a real-world example with real data!
    - Explore the functionality offered in the rdrobust package in R
    - Learn how to use the function rdrobust to estimate a causal effect