STA 235 - Causal Inference: Regression Discontinuity Design (Cont.)

Spring 2021

McCombs School of Business, UT Austin

Reminders

In-class midterm March 29th

- **What will the midterm look like?**
  - Shorter version of a homework: Examples/cases with data (conceptual + R code)
- **Where should I study from?**
  - R code posted for class (includes conceptual questions), R code example questions for midterm, examples/questions seen in class, JITTs, etc.
  - Other resources posted on the bookmark section (websites have *a lot* of data exercises)
  - **Don't memorize anything.**

Today

- Quick recap and finish with **regression discontinuity design**:
  - How do we estimate an effect in an RD?
- **Instrumental variables**:
  - Fuzzy regression discontinuity designs.

Let's recap

Behind the scenes of RDs

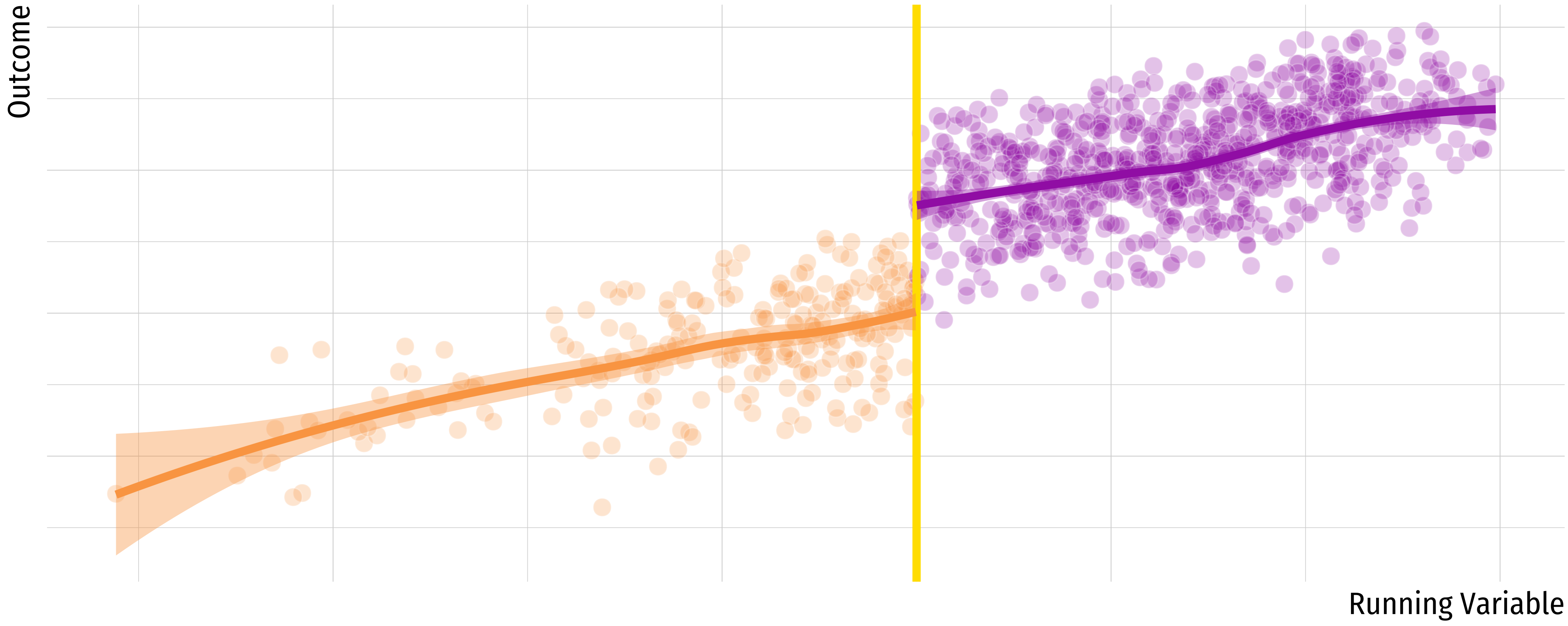
- Basically, regression discontinuities work under an **asymptotic assumption**:
- Let Y_i be the outcome of interest, Z_i the treatment assignment, R_i the running variable, and c the cutoff score:

$$Z_i = \begin{cases} 0 & R_i \leq c \\ 1 & R_i > c \end{cases}$$

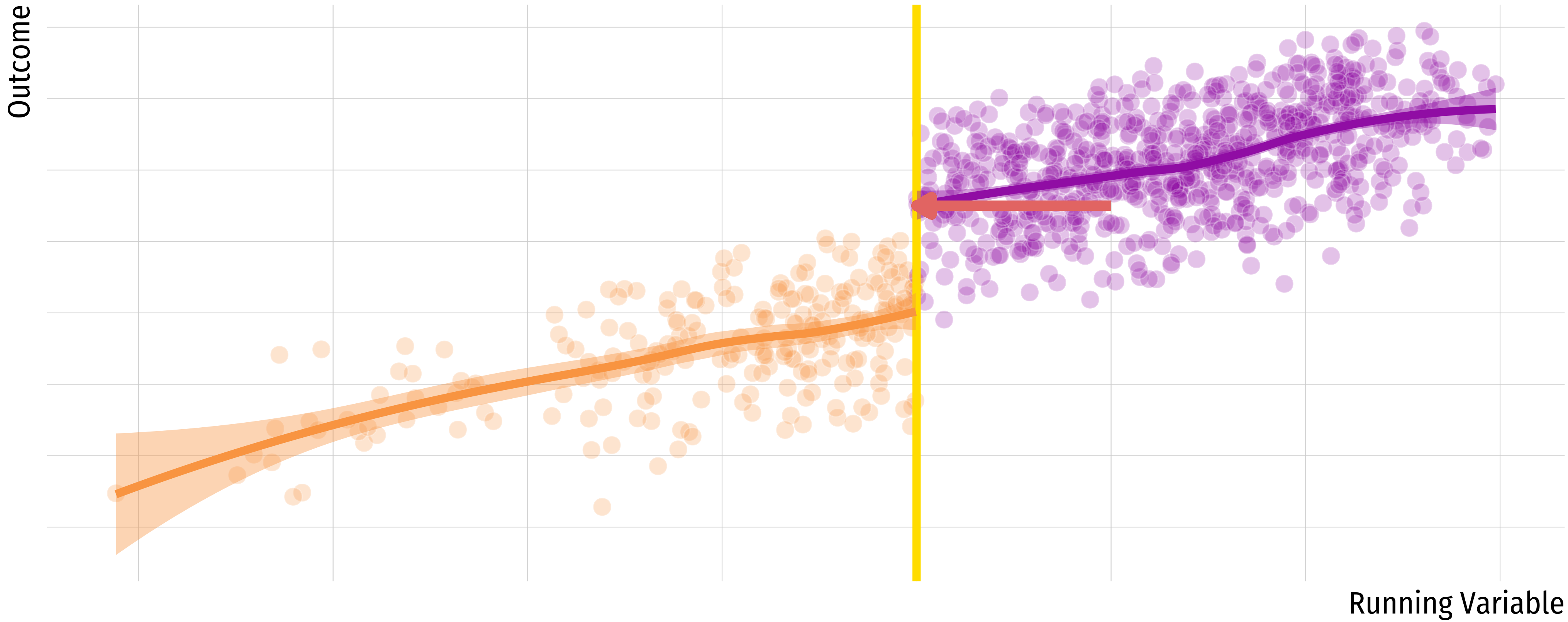
- Then, we can define the treatment effect δ as:

$$\delta = \lim_{\epsilon \to 0^+} E[Y_i | R_i = c + \epsilon] - \lim_{\epsilon \to 0^-} E[Y_i | R_i = c + \epsilon]$$

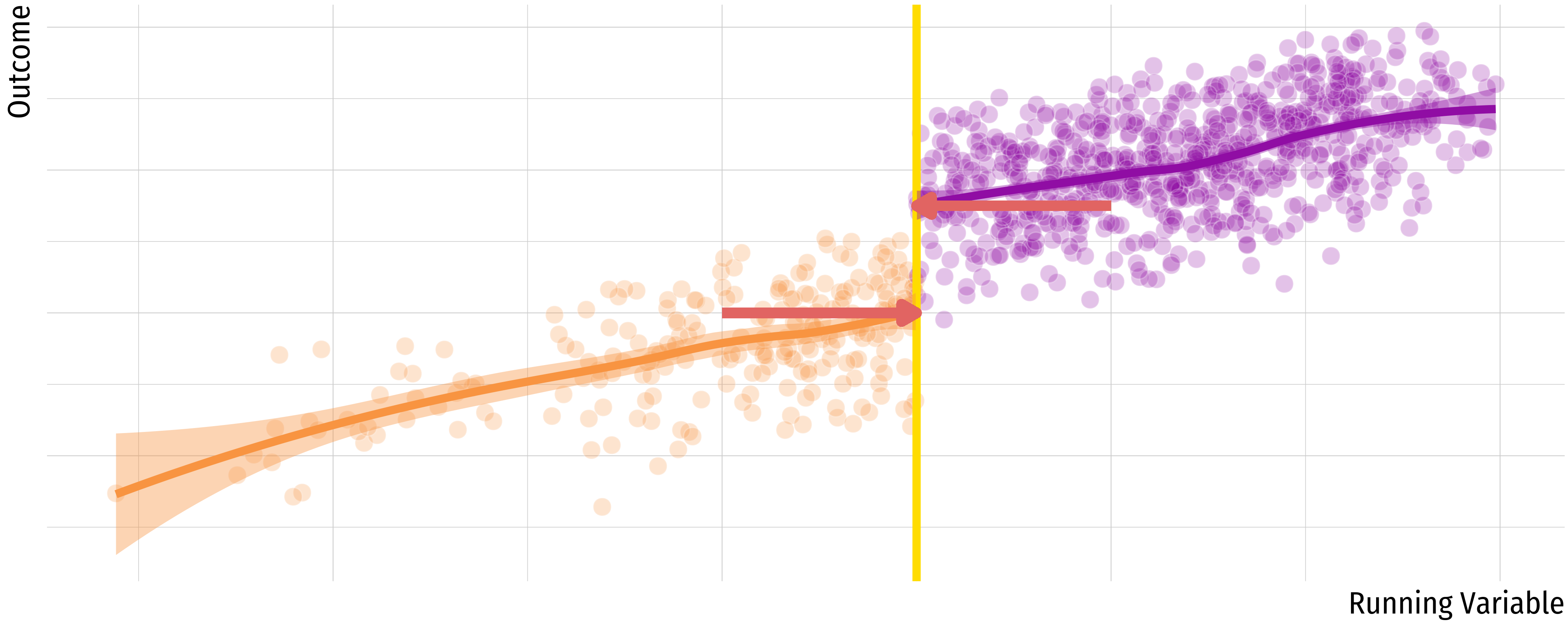
What does the limit expression mean?



What does the limit expression mean?



What does the limit expression mean?

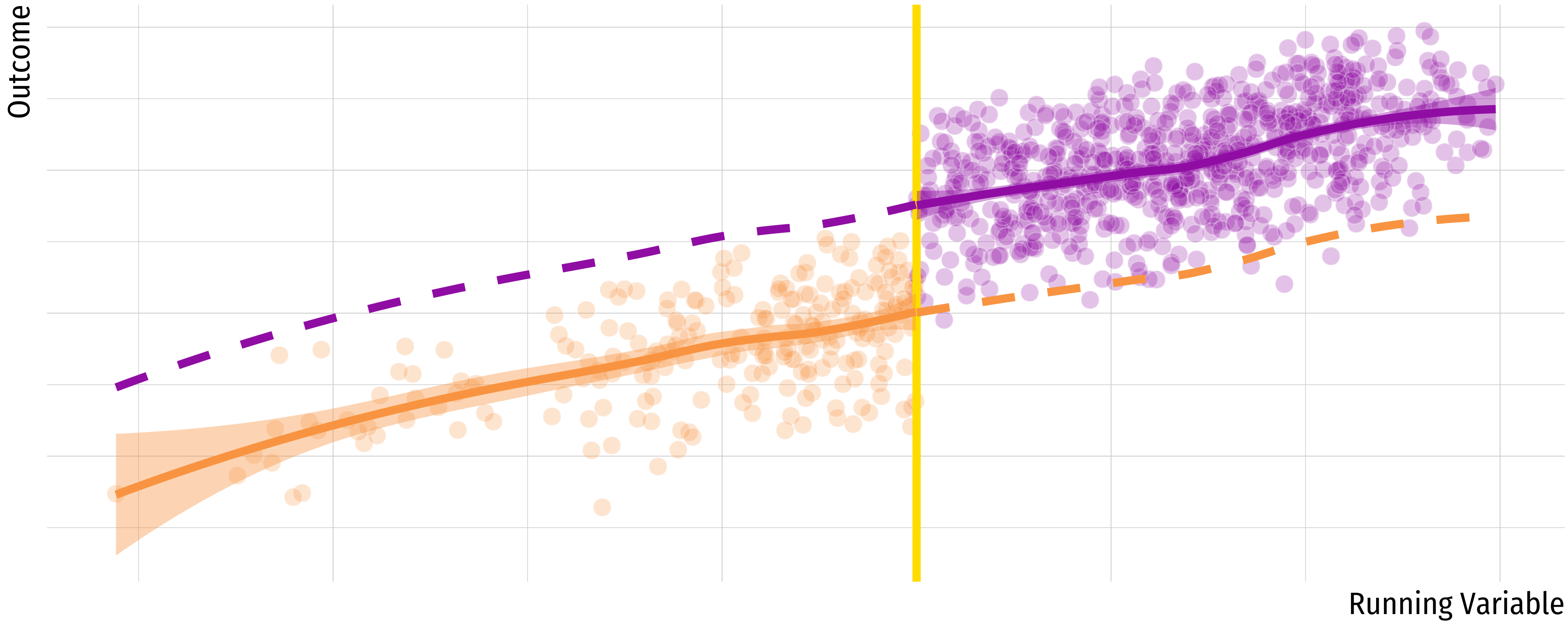


Conditions required for identification

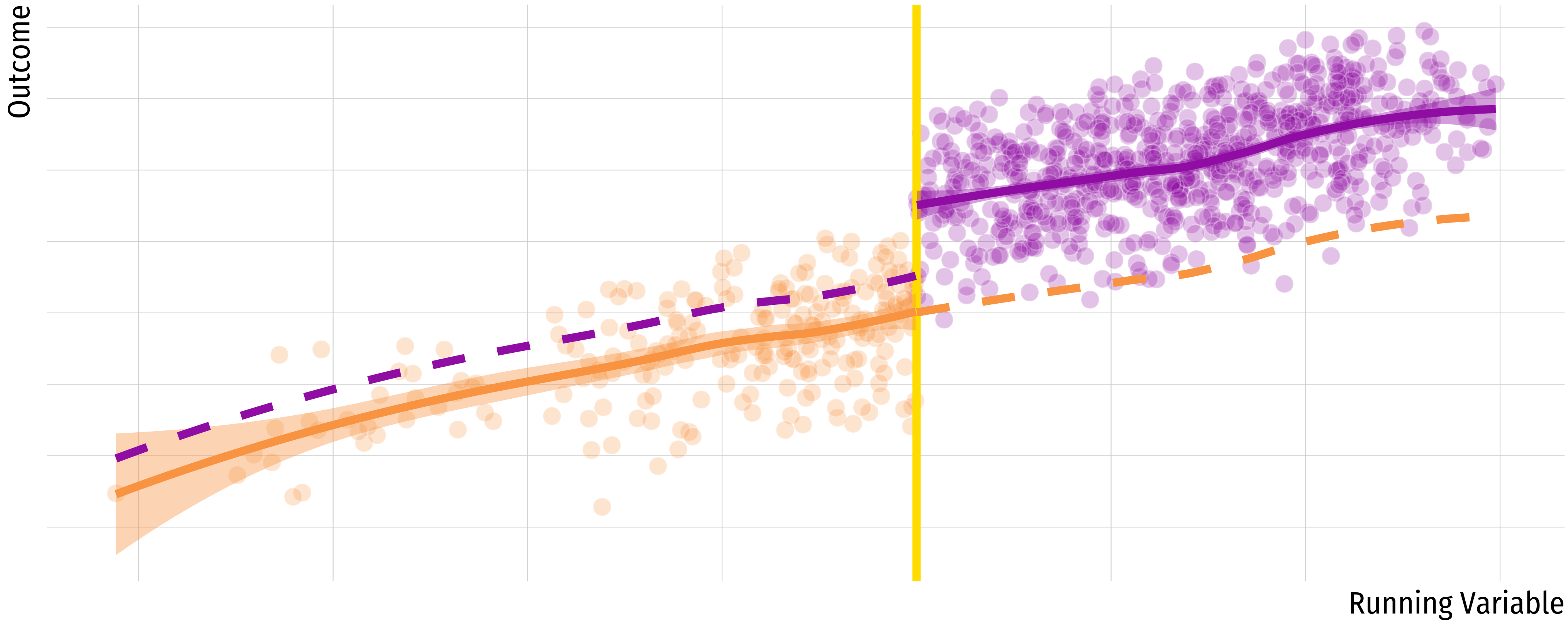
- Threshold rule **exists** and cutoff point is **known**
- The running variable R_i is **continuous** near c.
- Key assumption:

Continuity of E[Y(1)|R] and E[Y(0)|R] at R=c

Potential outcomes need to be smooth across the threshold



Potential outcomes need to be smooth across the threshold



How can I check if this assumption holds?

You can't! (it's an assumption)

Robustness checks:

- Check density across the cutoff
- Check RD for covariates

Estimation in practice

How do we actually estimate an RD?

- The simplest way to do this is to fit a regression:

$$Y_i = \beta_0 + \beta_1(R_i - c) + \beta_2 \mathrm{I}[R_i > c] + \beta_3(R_i - c)\mathrm{I}[R_i > c]$$

How do we actually estimate an RD?

- The simplest way to do this is to fit a regression:

$$Y_i = \beta_0 + \beta_1 \underbrace{(R_i - c)}_{\text{Distance to the cutoff}} + \beta_2 \mathbf{I}[R_i > c] + \beta_3 \overbrace{(R_i - c)}^{\text{Distance to the cutoff}} \mathbf{I}[R_i > c]$$

How do we actually estimate an RD?

- The simplest way to do this is to fit a regression:

$$Y_i = \beta_0 + \beta_1(R_i - c) + \beta_2\underbrace{\mathrm{I}[R_i > c]}_{\text{Treatment}} + \beta_3(R_i - c)\overbrace{\mathrm{I}[R_i > c]}^{\text{Treatment}}$$

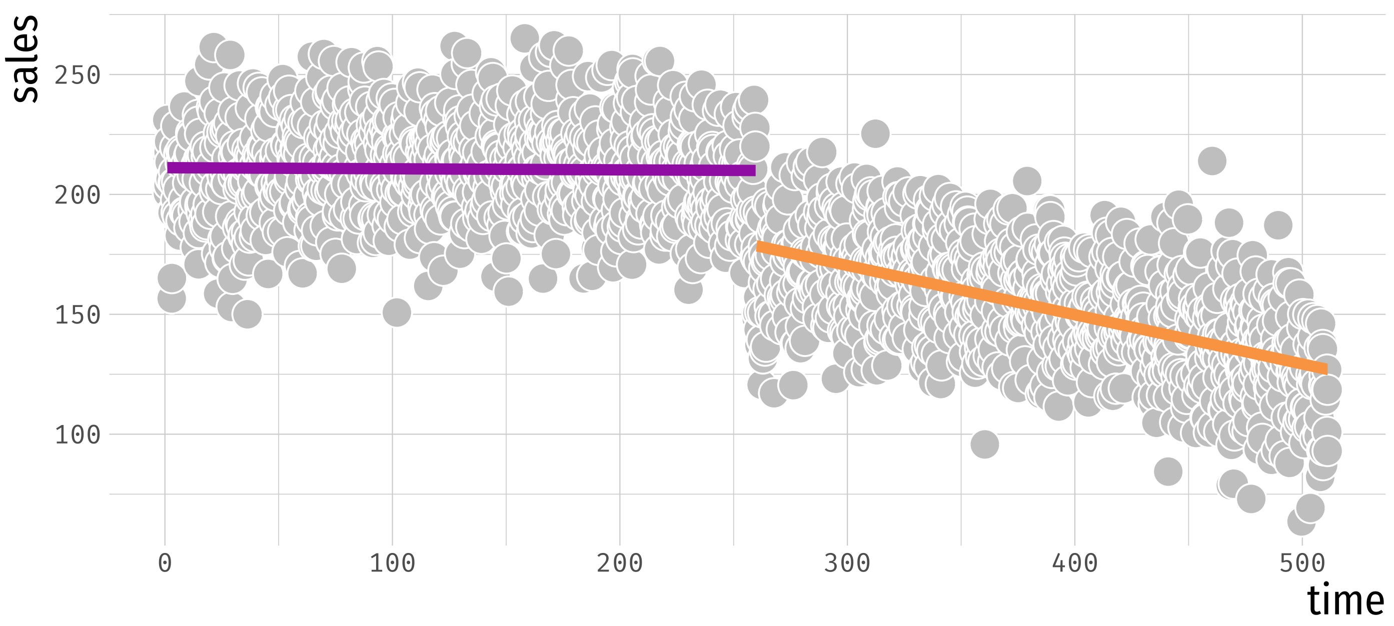
- You want to add **flexibility** for each side of the cutoff.

Can you identify these parameters in a plot?

Let's see some examples: Sales using a linear model

```
sales <- sales %>% mutate(dist = c-time)

lm(sales ~ dist + treat + dist*treat, data = sales)
```



Let's see some examples: Sales using a linear model

```
summary(lm(sales ~ dist + treat + dist*treat, data = sales))
```

```
##
## Call:
## lm(formula = sales ~ dist + treat + dist * treat, data = sales)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -65.738 -13.940   0.051  13.538  76.515
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 178.640954   1.300314  137.38   <2e-16 ***
## dist          0.205355   0.008882   23.12   <2e-16 ***
## treat        31.333952   1.842338   17.01   <2e-16 ***
## dist:treat   -0.200845   0.012438  -16.15   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.52 on 1996 degrees of freedom
## Multiple R-squared:  0.6939,    Adjusted R-squared:  0.6934
## F-statistic:  1508 on 3 and 1996 DF,  p-value: < 2.2e-16
```

We can be more flexible

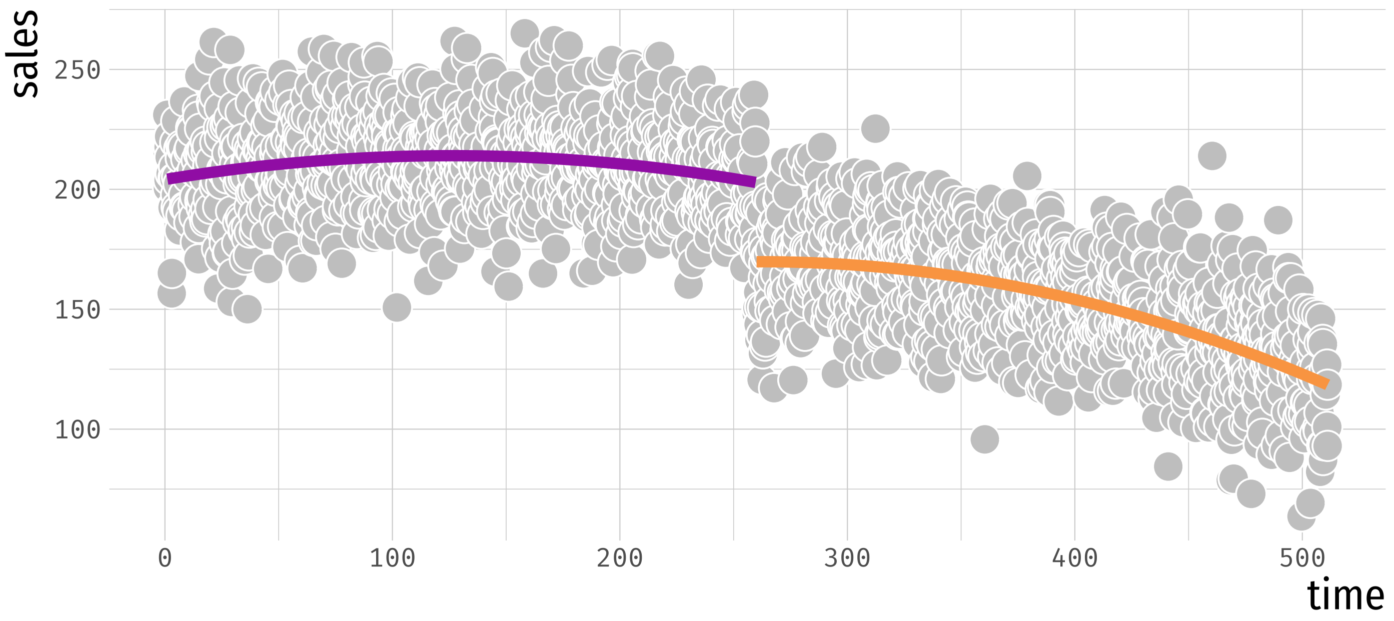
- The previous example just included linear terms, but you can also be more flexible:

$$Y_i = \beta_0 + \beta_1 f(R_i - c) + \beta_2 \mathrm{I}[R_i > c] + \beta_3 f(R_i - c)\mathrm{I}[R_i > c]$$

- Where f is any function you want.

What happens if we fit a quadratic model?

```
lm(sales ~ dist + I(dist^2) + treat + dist*treat + treat*I(dist^2), data = sales)
```



What happens if we fit a quadratic model?

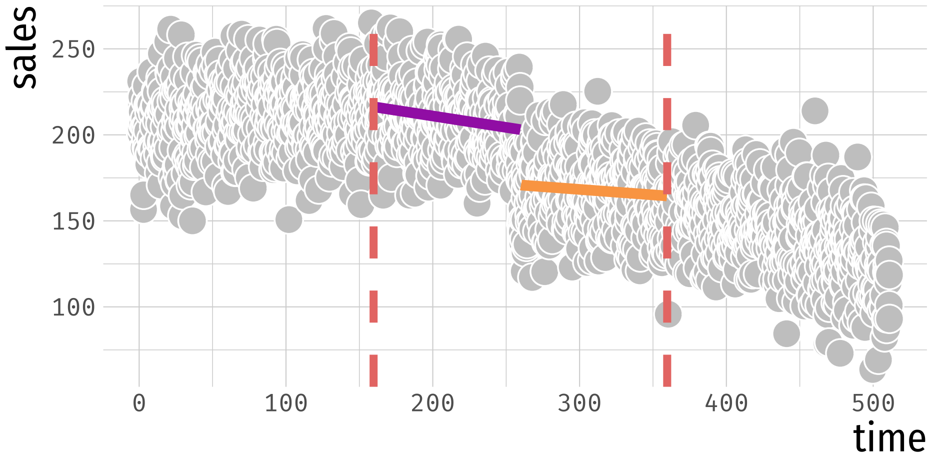
```
summary(lm(sales ~ dist + I(dist^2) + treat + dist*treat + treat*I(dist^2), data = sales))
```

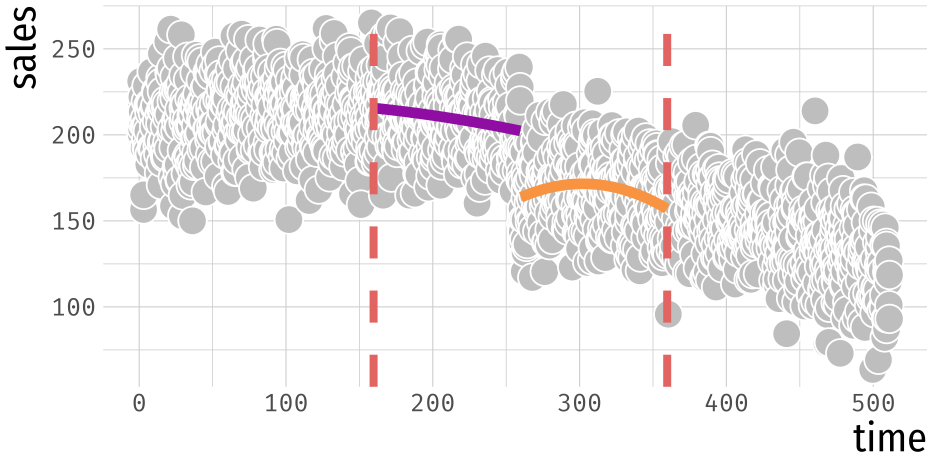
```
##
## Call:
## lm(formula = sales ~ dist + I(dist^2) + treat + dist * treat +
##     treat * I(dist^2), data = sales)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -66.090 -13.979   0.239  13.154  76.656
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.698e+02  1.937e+00  87.665  < 2e-16 ***
## dist            -4.302e-03  3.556e-02  -0.121 0.903725
## I(dist^2)       -8.288e-04  1.363e-04  -6.083 1.41e-09 ***
## treat            3.308e+01  2.747e+00  12.041  < 2e-16 ***
## dist:treat       1.713e-01  4.964e-02   3.452 0.000569 ***
## I(dist^2):treat  2.034e-04  1.877e-04   1.084 0.278554
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.23 on 1994 degrees of freedom
## Multiple R-squared:  0.7029,    Adjusted R-squared:  0.7021
## F-statistic: 943.5 on 5 and 1994 DF,  p-value: < 2.2e-16
```

What happens if we only look at observations close to c?

```
sales_close <- sales %>% filter(dist>-100 & dist<100)

lm(sales ~ dist + treat + dist*treat + treat, data = sales_close)
lm(sales ~ dist + I(dist^2) + treat + dist*treat + treat*I(dist^2), data = sales_close)
```





How do they compare?

```r
summary(lm(sales ~ dist + treat + dist*treat + treat, data = sales_close))
```

```
##
## Call:
## lm(formula = sales ~ dist + treat + dist * treat + treat, data = sales_close)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -53.241 -14.764   0.268  12.938  57.811
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 170.84457    2.05528  83.125   <2e-16 ***
## dist          0.06345    0.03542   1.791   0.0736 .
## treat        32.21243    2.93614  10.971   <2e-16 ***
## dist:treat    0.06909    0.05047   1.369   0.1714
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.25 on 782 degrees of freedom
## Multiple R-squared:  0.5261,    Adjusted R-squared:  0.5243
## F-statistic: 289.4 on 3 and 782 DF,  p-value: < 2.2e-16
```

How do they compare?

```r
summary(lm(sales ~ dist + I(dist^2) + treat + dist*treat + treat*I(dist^2), data = sales_close))
```

```
##
## Call:
## lm(formula = sales ~ dist + I(dist^2) + treat + dist * treat +
##     treat * I(dist^2), data = sales_close)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -50.080 -14.238  -0.463  12.740  54.231
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)     163.550012   3.001833  54.483  < 2e-16 ***
## dist             -0.375526   0.136936  -2.742 0.006240 **
## I(dist^2)        -0.004415   0.001331  -3.317 0.000951 ***
## treat            38.757140   4.316684   8.978  < 2e-16 ***
## dist:treat        0.552254   0.195847   2.820 0.004927 **
## I(dist^2):treat   0.003975   0.001894   2.099 0.036121 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.13 on 780 degrees of freedom
## Multiple R-squared:  0.5328,    Adjusted R-squared:  0.5298
## F-statistic: 177.9 on 5 and 780 DF,  p-value: < 2.2e-16
```

Potential problems

- There are **many potential problems** with the previous examples:
    - Which polynomial function should we choose? Linear, quadratic, other?
    - What bandwidth should we choose? Whole sample? [-100,100]?

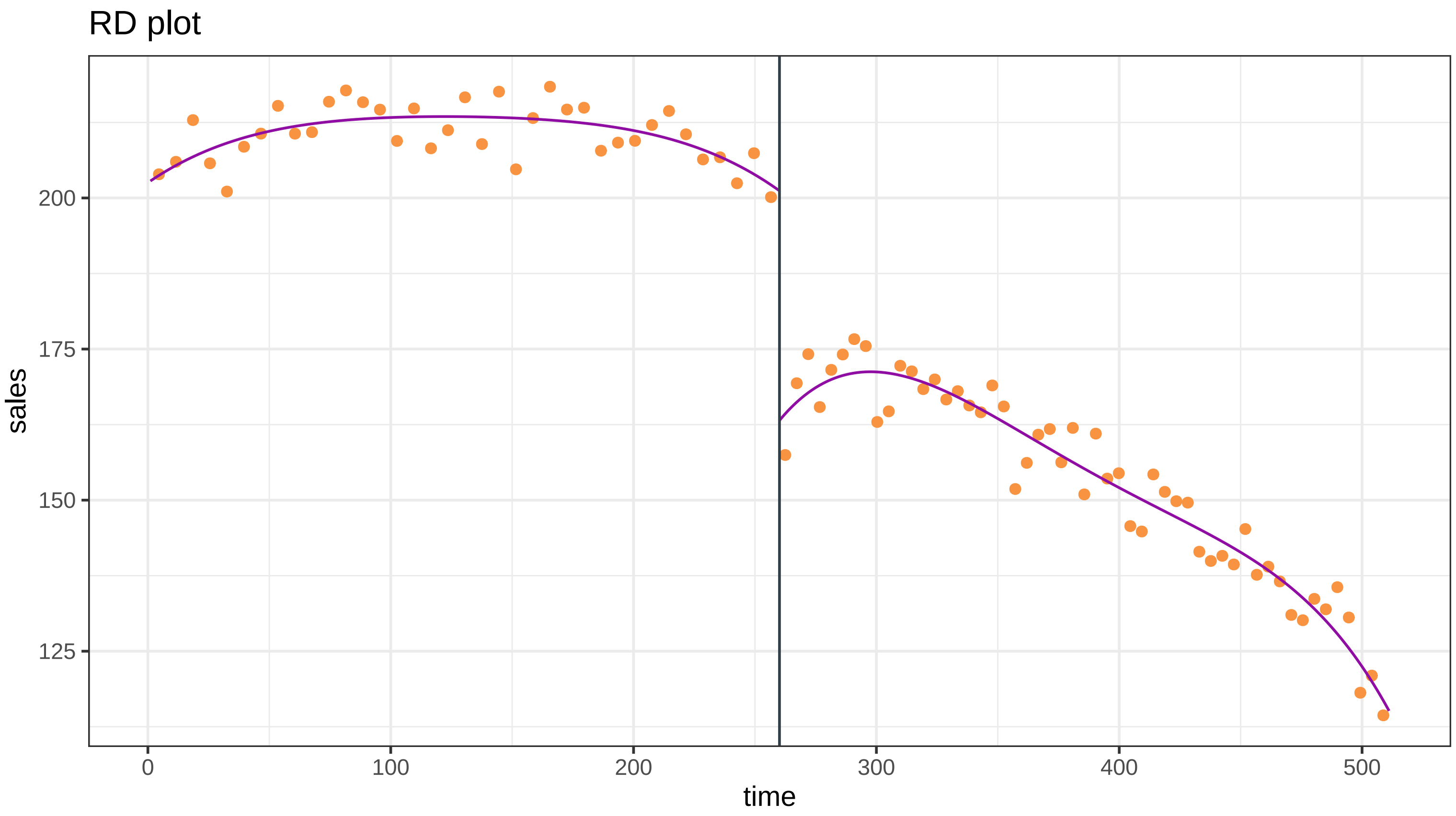
- There are some ways to address these concerns.

Package `rdrobust`

- Robust Regression Discontinuity introduced by Cattaneo, Calonico, Farrell & Titiunik (2014).
- Use of **local polynomial** for fit.
- **Data-driven optimal bandwidth** (bias vs variance).
- `rdrobust`: Estimation of LATE and opt. bandwidth
- `rdplot`: Plotting RD with nonparametric local polynomial.

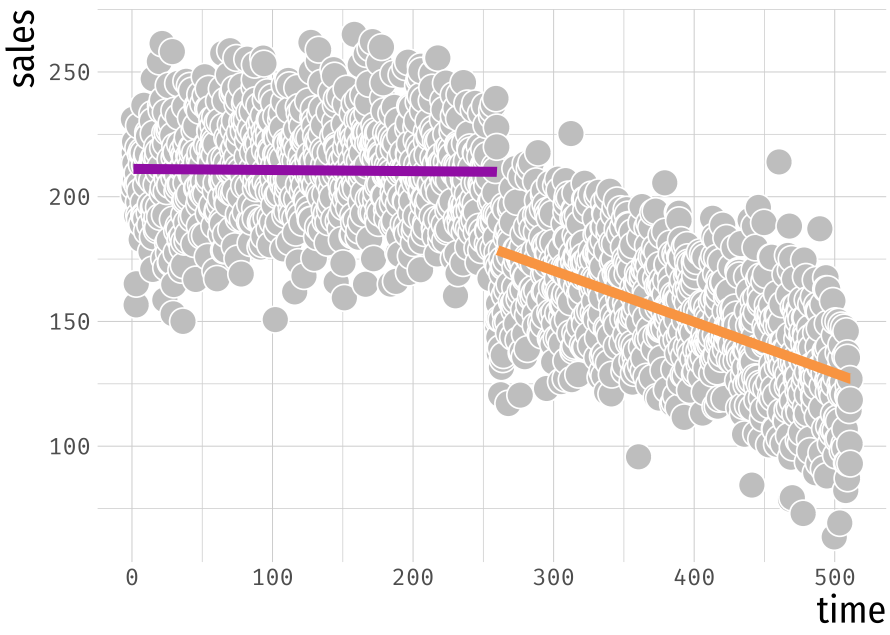
Let's compare with previous parametric results

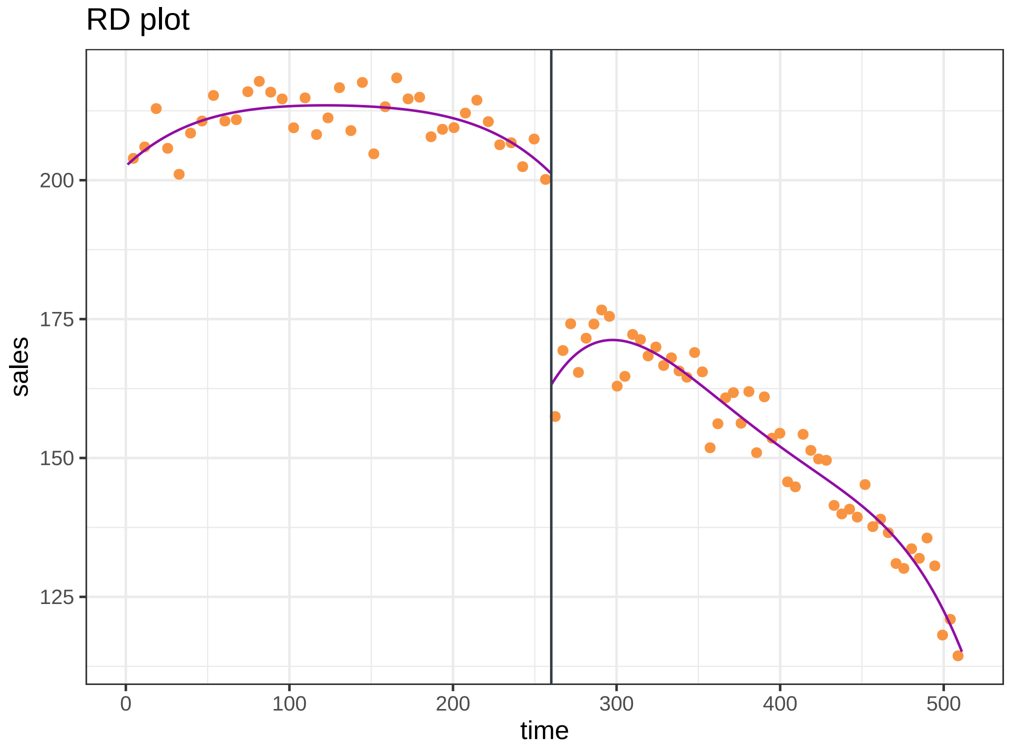
```r
rdplot(y = sales$sales, x = sales$time, c = c,
       title = "RD plot", x.label = "time", y.label = "sales")
```



Let's compare with previous parametric results

```
rdplot(y = sales$sales, x = sales$time, c = c,
       title = "RD plot", x.label = "time", y.label = "sales")
```





Let's compare with previous parametric results

```
summary(rdrobust(y = sales$sales, x = sales$time, c = c))
```

```
## Call: rdrobust
##
## Number of Obs.                  2000
## BW type                        mserd
## Kernel                    Triangular
## VCE method                        NN
##
## Number of Obs.                 1000         1000
## Eff. Number of Obs.             202          213
## Order est. (p)                    1            1
## Order bias  (q)                   2            2
## BW est. (h)                  54.304       54.304
## BW bias (b)                  87.787       87.787
## rho (h/b)                     0.619        0.619
## Unique Obs.                    1000         1000
##
## =============================================================================
##         Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## =============================================================================
##   Conventional   -37.434     4.344    -8.618     0.000   [-45.948 , -28.921]
##         Robust         -         -    -7.610     0.000   [-48.596 , -28.691]
## =============================================================================
```

How do we weight observations?

- `rdrobust` uses `rdbwselect()` function (by default) to estimate a data-driven bandwidth (i.e. what observations we are going to use for estimation).
  - If we use a bandwidth, does this mean that the RD is estimating an effect for that population within the bandwidth?
- **Kernels** are also important in this context:
  - How do I weight observations within the bandwidth (e.g. uniform, triangle)

Observing kernels

Takeaway points

- RD designs are **great** for causal inference!
  - Strong internal validity
  - Number of robustness checks
- **Limited** external validity.
- Make sure to check your data:
  - Discontinuity in treatment assignment
  - Density across the cutoff
  - Smoothness of covariates

References

- Angrist, J. and S. Pischke. (2015). "Mastering Metrics". *Chapter 4*.
- Calonico, Cattaneo and Titiunik. (2015). "rdrobust: An R Package for Robust Nonparametric Inference in Regression-Discontinuity Designs". R Journal 7(1): 38-51.
- Heiss, A. (2020). "Program Evaluation for Public Policy". *Class 10: Regression Discontinuity I, Course at BYU*.
- Lee, D. and T. Lemieux. (2010). "Regression Discontinuity in Economics". *Journal of Economic Literature 48, pp 281-355*.