• What is regression discontinuity
  • Background
  • Definition
  • Uses in educational research
  • Design, analysis, model
  • Assumptions and considerations

• Regression discontinuity analysis in R
  • Example 1: No statistical significance
  • Example 2: Main effect
  • Example 3: Interaction effect
  • Example 4: Main and interaction effect
  • Example 5: Multiple groups
Background

• Brought to forefront by Thistlethwaite and Campbell (1960)
  • Mimicking effects of randomized selection without bias
  • Developing methods for statistical analysis

• Used within other fields of research, such as medicine and economics, before popularity grew in education. (Cook, 2008)
Definition

- Quasi-experimental research design
  - Unique assignment method
  - Cut-off criterion to select groups
    - Pre-determined
    - Completely known

- An alternative to randomized control trial experiments
  - Assign participants to groups based on an observed variable
    - Ex. Standardized assessment score, GPA

- A method used for program evaluation
  - Compare outcomes of treatment group determined by cutoff criterion to outcomes of comparison group
Education Research

**Treatment Assignment**

- Skipping class in college and exam performance: An RD class experiment (Dobkin, Gil, & Marion, 2009)
  - Students below median on midterm assigned mandatory attendance policy.
  - Students above median were not.
  - Results on final exam of two groups analyzed.

**Program Evaluation**

- Impact of Texas top 10% law on college enrollment (Niu & Tienda, 2010)
  - Studied minority enrollment trends at UT and TAMU as predicted by class rank
  - Comparison group were students ranking at or below 10%.
  - Treatment group were students ranking above 10%.
Design

Below cut-off group

Above cut-off group

Outcome Variable

Assignment Variable (ordinal)

Cut off criterion

Does a discontinuity occur in the regression at the cutoff point?
• Does a discontinuity occur in the regression at the cutoff point?

Main Effect
Change in intercept

Interaction Effect
Change in slope

Main & Interaction Effect
Change in intercept & slope
Design: Skipping Class Study

Did class attendance impact exam scores?
Design: Texas Top 10% Law

How did the law impact minority enrollment rates?
Model

\[ y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + e_i \]
Assumptions

- **Cutoff Criterion**: Must be followed without exception
- **Pre-post distribution**: Must not be better explained with a curve
- **Comparison group pretest variance**: Must have sufficient number of values
- **Continuous pretest distribution**: Both groups came from same distribution
- **Equal treatment**: Treatment administered in same way for all participants
Considerations

**Strengths**
- Unbiased assignation
- Unbiased estimate of effects of intervention
- Good alternative to RCT
- Practical method for school settings

**Limitations**
- Need to meet many assumptions (fuzzy designs)
- Need to model results correctly
- Need ~3x sample size of RCT
- Not as statistically powerful
- Discontinuity could be curvilinear relationship
Curvilinearity

- When the discontinuity is better explained by a curvilinear relationship instead of regression lines.

- Best way to avoid this—specify model correctly.
library (lattice)

1. Transform the pre-treatment value (x).
   - Subtract cutoff score from pre-treatment score.
     \[ x_T = x_i - x_c \]

2. Dummy code the assignment variable.
   - 0 = pre-treatment, 1 = post-treatment

3. Run a regression on dependent variable with transformed x and dummy variable.
library (lattice)

DATA SET 1

group1<-data.frame(time1=1:20,score1=(c(10,15,20,25,30,35,40,45,50,55,60,65,70,75,80,85,90,95,100,105)),interv1=(factor(rep(0:1,each=10))))
group1$score1<-jitter(group1$score1,factor=12)

DETERMINE CUTOFF CRITERION. (10.5)

VISUAL ANALYSIS

xyplot(score1~time1,group1,xlab="Time",ylab="Score",main="Progression through Intervention",pch=c(19,17),groups=interv1,col=c("blue","green"),type=c("p","r"),lwd=2,lty=c(2,1))

STATISTICAL ANALYSIS

analysis1<-lm(score1~I(time1-10.5)*interv1,group1)
summary(analysis1)
**Example 1: No statistical significance**

- **R Code**
  ```r
  analysis1<-lm(score1~I(time1-10.5)*interv1,group1)
  > summary(analysis1)
  
  Call:
  lm(formula = score1 ~ I(time1 - 10.5) * interv1, data = group1)
  
  Residuals:
  Min      1Q  Median      3Q     Max
  -9.5498 -3.5464 -0.2439  3.2496  8.4917

  Coefficients:
  Estimate  Std. Error   t value  Pr(>|t|)
  (Intercept)                50.7682     3.8568  13.163 5.33e-10 ***
  I(time1 - 10.5)            3.4603     0.6689   5.173 9.24e-05 ***
  interv11                   5.4994     5.4544   1.008    0.328
  I(time1 - 10.5):interv11   1.8591     0.9459   1.965    0.067 .

  ---
  Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

  Residual standard error: 6.075 on 16 degrees of freedom
  Multiple R-squared: 0.9634,    Adjusted R-squared: 0.9565
  F-statistic: 140.2 on 3 and 16 DF,  p-value: 1.069e-11
  ```
Example 1: No statistical significance

\[ y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \beta_3 x_i z_i + e_i \]

\[ y = 50.77 + 3.46(time1-10.5) + 5.50(interv1) + 1.86(time-10.5)(interv1) \]

\[ \text{PRE-CUTOFF GROUP} \]
\[ y = 50.77 + 3.46(time1-10.5) + 5.50(0) + 1.86(time1-10.5)(0) \]
\[ y = 50.77 + 3.46(time1-10.5) \]

\[ \text{POST-CUTOFF GROUP} \]
\[ y = 50.77 + 3.46(time1-10.5) + 5.50(1) + 1.86(time1-10.5)(1) \]
\[ y = 56.27 + 5.32(time1-10.5) \]
library (lattice)

**DATA SET 2**
group2<-data.frame(time2=1:20,score2=c(5,10,15,20,25,30,35,40,45,50,100,105,110,115,120,125,130,135,140,145),interv2=(factor(rep(0:1,each=10))))
group2$score2<-jitter(group2$score2,factor=10)

DETERMINE CUTOFF CRITERION. (10.5)

**VISUAL ANALYSIS**

xyplot(score2~time2,group2,xlab="Time",ylab="Score",main="Progression through Intervention Scenario 2",pch=c(19,17),groups=interv2,col=c("blue","red"),type=c("p","r"),lwd=2,lty=c(2,1))

**STATISTICAL ANALYSIS**

analysis2<-lm(score2~I(time2-10.5)*interv2,group2)
summary(analysis2)
Example 2: Main Effect

```
> analysis2 <- lm(score2 ~ I(time2 - 10.5) * interv2, group2)
> summary(analysis2)

Call:
  lm(formula = score2 ~ I(time2 - 10.5) * interv2, data = group2)

Residuals:      
   Min       1Q   Median       3Q      Max
-8.8572 -4.5041 -0.1747  4.3713 12.5243

Coefficients:       Estimate Std. Error t value Pr(>|t|)
  (Intercept)         55.0152     4.1225  13.345 4.35e-10 ***
  I(time2 - 10.5)    5.6650     0.7149   7.924 6.28e-07 ***
  interv21           44.2711     5.8301   7.594 1.08e-06 ***
  I(time2 - 10.5):interv21 -1.0723     1.0111  -1.061    0.305

---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 . ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.494 on 16 degrees of freedom
Multiple R-squared: 0.9867,   Adjusted R-squared: 0.9842
F-statistic: 395.6 on 3 and 16 DF,  p-value: 3.253e-15
```
Example 2: Main Effect

\[ y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \beta_3 x_i z_i + e_i \]

\[ y = 55.02 + 5.67(time2-10.5) + 44.27(interv2) - 1.07(time2-10.5)(interv2) \]

PRE-CUTOFF GROUP
\[ y = 55.02 + 5.67(time2-10.5) + 44.27(0) - 1.07(time2-10.5)(0) \]
\[ y = 55.02 + 5.67(time2-10.5) \]

POST-CUTOFF GROUP
\[ y = 55.02 + 5.67(time2-10.5) + 44.27(1) - 1.07(time2-10.5)(1) \]
\[ y = 99.29 + 4.60(time2-10.5) \]
library (lattice)

**DATA SET 3**
group3<-data.frame(time3=1:20,score3=c(50:59,60,65,70,75,80,85,90,95,100,105),interv3=(factor(rep(0:1,each=10))))
group3$score3<-jitter(group3$score3,factor=10)

DETERMINE CUTOFF CRITERION. (10.5)

**VISUAL ANALYSIS**

xyplot(score3~time3,group3,xlab="Time",ylab="Score",
main="Progression through Intervention Scenario 3",pch=c(19,17),
groups=interv3,col=c("maroon","grey"),type=c("p","r"),lwd=2,lty=c(2,1))

**STATISTICAL ANALYSIS**

analysis3<-lm(score3~I(time3-10.5)*interv3,group3)
summary(analysis3)
Example 3: Interaction Effect

```
analysis3 <- lm(score3 ~ I(time3 - 10.5) * interv3, group3)
> summary(analysis3)

Call:
  lm(formula = score3 ~ I(time3 - 10.5) * interv3, data = group3)

Residuals:
     Min       1Q   Median       3Q      Max
-1.7942 -0.6819 -0.1268  0.8217  1.7521

Coefficients:                           Estimate Std. Error t value Pr(>|t|)
(Intercept)                60.4268     0.6981  86.564  < 2e-16 ***
I(time3 - 10.5)            1.1573     0.1211   9.560 5.12e-08 ***
interv31                  -1.9138     0.9872  -1.939   0.0704 .
I(time3 - 10.5):interv31   3.6504     0.1712  21.322 3.56e-13 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 . ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.1 on 16 degrees of freedom
Multiple R-squared: 0.9967,   Adjusted R-squared: 0.9961
F-statistic: 1630 on 3 and 16 DF,  p-value: < 2.2e-16
```
Example 3: Interaction Effect

\[ y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \beta_3 x_i z_i + e_i \]

\[ y = 60.43 + 1.16 (\text{time3-10.5}) - 1.91 (\text{interv3}) + 3.65 (\text{time3-10.5})(\text{interv3}) \]

```
analysis3<-lm(score3~I(time3-10.5)*interv3,group3)
> summary(analysis3)

Call:
  lm(formula = score3 ~ I(time3 - 10.5) * interv3, data = group3)

Residuals:
               Min          1Q      Median          3Q         Max
-1.7942      -0.6819      -0.1268       0.8217      1.7521

Coefficients:
                              Estimate Std. Error  t value Pr(>|t|)
(Intercept)                     60.4268     0.6981   86.564  < 2e-16 ***
I(time3 - 10.5)                 1.1573     0.1211    9.560   5.12e-08 ***
interv31                       -1.9138     0.9872   -1.939    0.0704 .
I(time3 - 10.5):interv31        3.6504     0.1712   21.322  3.56e-13 ***

---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.1 on 16 degrees of freedom
Multiple R-squared: 0.9967, Adjusted R-squared: 0.9961
F-statistic: 1630 on 3 and 16 DF,  p-value: < 2.2e-16
```

**PRE-CUTOFF GROUP**

\[ y = 60.43 + 1.16 (\text{time3-10.5}) - 1.91(0) + 3.65(\text{time3-10.5})(0) \]

\[ y = 60.43 + 1.16 (\text{time3-10.5}) \]

**POST-CUTOFF GROUP**

\[ y = 60.43 + 1.16 (\text{time3-10.5}) - 1.91(1) + 3.65(\text{time3-10.5})(1) \]

\[ y = 58.52 + 4.81 (\text{time3-10.5}) \]
Example 4: Main & Interaction Effect

```r
library (lattice)

DATA SET 4

group4<-data.frame(time4=1:20,score4=c(50:59, 80,85,90,95,100,105,110,115,120,125),interv4=(factor(rep(0:1,each=10))))
group4$score4<-jitter(group4$score4,factor=12)

DETERMINE CUTOFF CRITERION. (10.5)

VISUAL ANALYSIS

xyplot(score4~time4,group4,xlab="Time",ylab="Score",
main="Progression through Intervention Scenario 4",pch=c(19,17),
groups=interv4,col=c("blue","grey"),type=c("p","r"),lwd=2,lty=c(2,1))

STATISTICAL ANALYSIS

analysis4<-lm(score4~I(time4-10.5)*interv4,group4)
summary(analysis4)
```
Example 4: Main & Interaction Effect

```r
analysis4<-lm(score4~I(time4-10.5)*interv4,group4)
> summary(analysis4)

Call:
  lm(formula = score4 ~ I(time4 - 10.5) * interv4, data = group4)

Residuals:
     Min       1Q   Median       3Q      Max
-1.8244 -0.5385  0.2426  0.5778  1.9935

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)            58.5511     0.6854  85.432  < 2e-16 ***
I(time4 - 10.5)        0.8715     0.1189   7.332  1.68e-06 ***
interv41              17.7430     0.9692  18.306  3.73e-12 ***
I(time4 - 10.5):interv41  4.3282     0.1681  25.750  1.89e-14 ***

---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.08 on 16 degrees of freedom
Multiple R-squared: 0.9987,    Adjusted R-squared: 0.9984
F-statistic: 3964 on 3 and 16 DF,  p-value: < 2.2e-16
```
Example 4: Main & Interaction Effect

\[ y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \beta_3 x_i z_i + e_i \]

\[ y = 58.55 + 0.87(time4-10.5) + 17.74(interv4) + 4.33(time4-10.5)(interv4) \]

PRE-CUTOFF GROUP

\[ y = 58.55 + 0.87(time4-10.5) + 17.74(0) + 4.33(time4-10.5)(0) \]

\[ y = 58.55 + 0.87 (time4-10.5) \]

POST-CUTOFF GROUP

\[ y = 58.55 + 0.87(time4-10.5) + 17.74(1) + 4.33(time4-10.5)(1) \]

\[ y = 76.29 + 5.20 (time4-10.5) \]
library (lattice)

**DATA SET 5**

group5<-data.frame(time5=1:30,score5=c(50:59, 60,65,70,75,80,85,90,95,100,105,106:115),interv5=(factor(rep(0:2,each=10))))
group5

group5$score5<-jitter(group5$score5,factor=12)

DETERMINE CUTOFF CRITERION. (10.5)

**VISUAL ANALYSIS**

xyplot(score5~time5,group5,xlab="Time",ylab="Score",
main="Progression through Intervention Scenario 5",pch=c(19,17),groups=interv5,
col=c("blue","green","gold"),type=c("p","r"),lwd=2,lty=c(2,1))

**STATISTICAL ANALYSIS**

analysis5<-lm(score5~I(time5-10.5)*interv5,group5)
summary(analysis5)
Example 5: Multiple Groups

Example 5: Multiple Groups

\[ \text{analysis5}<-\text{lm(score5}\sim\text{I(time5-10.5)\times interv5,group5)} \]
\[ > \text{summary(analysis5)} \]

Call:
\[ \text{lm(formula = score5 \sim I(time5 - 10.5) \times interv5, data = group5)} \]

Residuals:

Min       1Q  Median       3Q      Max
-2.3839 -1.2123 -0.4481  1.4422  2.3807

Coefficients:

|              | Estimate | Std. Error | t value | Pr(>|t|) |
|--------------|----------|------------|---------|----------|
| (Intercept)  | 60.0638  | 1.0027     | 59.901  | < 2e-16  *** |
| I(time5 - 10.5) | 1.0397   | 0.1739     | 5.979   | 3.59e-06 *** |
| interv51     | -1.9882  | 1.4181     | -1.402  | 0.174    |
| interv52     | 39.3363  | 2.8388     | 13.857  | 6.02e-13 *** |
| I(time5 - 10.5):interv51 | 3.8519   | 0.2459     | 15.663  | 4.22e-14 *** |
| I(time5 - 10.5):interv52 | -0.2932  | 0.2459     | -1.192  | 0.245    |

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.579 on 24 degrees of freedom
Multiple R-squared: 0.9966,    Adjusted R-squared: 0.9959
F-statistic: 1414 on 5 and 24 DF,  p-value: < 2.2e-16
Example 5: Multiple Groups

\[ y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \beta_3 z_i + \beta_4 x_i z_i + \beta_5 x_i z_i + e_i \]

\[ y = 60.06 + 1.04 (\text{time5} - 10.5) - 1.99 (\text{interv5}) + 39.34 (\text{interv5}) + 3.85 (\text{time5} - 10.5)(\text{interv5}) - 0.29 (\text{time5} - 10.5)(\text{interv5}) \]

\[
\text{analysis5} <- \text{lm(score5 ~ I(time5 - 10.5) * interv5, group5)}
\]

\text{summary(analysis5)}

Call:
\text{lm(formula = score5 ~ I(time5 - 10.5) * interv5, data = group5)}

Residuals:
Min 1Q Median 3Q Max
-2.3839 -1.2123 -0.4481 1.4422 2.3807

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 60.0638 1.0027 59.901 < 2e-16 ***
I(time5 - 10.5) 1.0397 0.1739 5.979 3.59e-06 ***
interv51 -1.9882 1.4181 -1.402 0.174
interv52 39.3363 2.8388 13.857 6.02e-13 ***
I(time5 - 10.5):interv51 3.8519 0.2459 15.663 4.22e-14 ***
I(time5 - 10.5):interv52 -0.2932 0.2459 -1.192 0.245
---
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.579 on 24 degrees of freedom
Multiple R-squared: 0.9966,    Adjusted R-squared: 0.9959
F-statistic: 1414 on 5 and 24 DF,  p-value: < 2.2e-16

PRE-CUTOFF GROUP

\[ y = 60.06 + 1.04 (\text{time5} - 10.5) - 1.99 (0) + 39.34 (0) + 3.85 (\text{time5} - 10.5)(0) - 0.29 (\text{time5} - 10.5)(0) \]

\[ y = 60.06 + 1.04 (\text{time5} - 10.5) \]

POST-CUTOFF GROUP 1

\[ y = 60.06 + 1.04 (\text{time5} - 10.5) - 1.99 (1) + 39.34 (1) + 3.85 (\text{time5} - 10.5)(1) - 0.29 (\text{time5} - 10.5)(1) \]

\[ y = 97.41 + 4.60 (\text{time5} - 10.5) \]

POST-CUTOFF GROUP 2

\[ y = 60.06 + 1.04 (\text{time5} - 10.5) - 1.99 (2) + 39.34 (2) + 3.85 (\text{time5} - 10.5)(2) - 0.29 (\text{time5} - 10.5)(2) \]

\[ y = 134.76 + 8.16 (\text{time5} - 10.5) \]
Conclusion

- Good option when unable to do RCT, especially in educational settings.
- Try to use as large a sample size as possible.
- Analyze data visually and statistically.
- Meet all assumptions—sharp is better than fuzzy.
- Beware of the curve.
- Specify your model correctly, and err on the side of overspecifying.


