Regression discontinuity

You work in a school district that is considering adding a special honors program for students that score well on a pretest at the beginning of high school. Students who score 80 or higher on the exam are automatically enrolled in honors classes that provide students with extra enrichment activities and outside-of-class support and tutoring. Administrators hope that the honors program will increase students' final GPA.

You have access to observational (i.e. not experimental) administrative data on students at the school, with the following columns:

Variable name	Description
`id`	Student ID number
`pretest`	Pretest score
`honors`	Indicator for being in the honors program
`gpa`	Final high school GPA

Your colleague attempted to measure the causal effect of this honors program on final GPA. They conducted some statistical analysis in R, but they forgot to interpret anything in the document, and now they've moved to a different district!

Given the information provided below, interpret the results from this analysis, as well as any assumption checks or tests your colleague included. Did this honors program have have an effect on GPAs? How much? Is it significant?

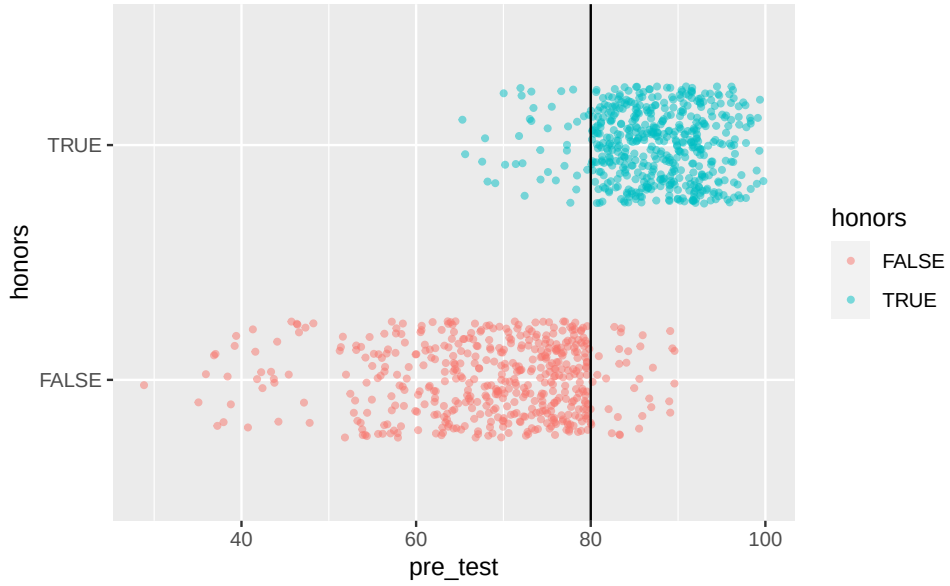
```
library(tidyverse)
library(broom)
library(rdrobust)
library(rddensity)
library(estimatr)

program <- read_csv("honors_program_data.csv") %>%
  mutate(pre_test_centered = pre_test - 80)
```

```
head(program)
```

id	pre_test	gpa	honors	pre_test_centered
1	92.4	3.78	TRUE	12.41
2	72.8	2.85	FALSE	-7.23
3	53.7	2.69	FALSE	-26.35
4	98.3	3.38	TRUE	18.33
5	69.7	2.35	FALSE	-10.29
6	68.1	2.57	FALSE	-11.93

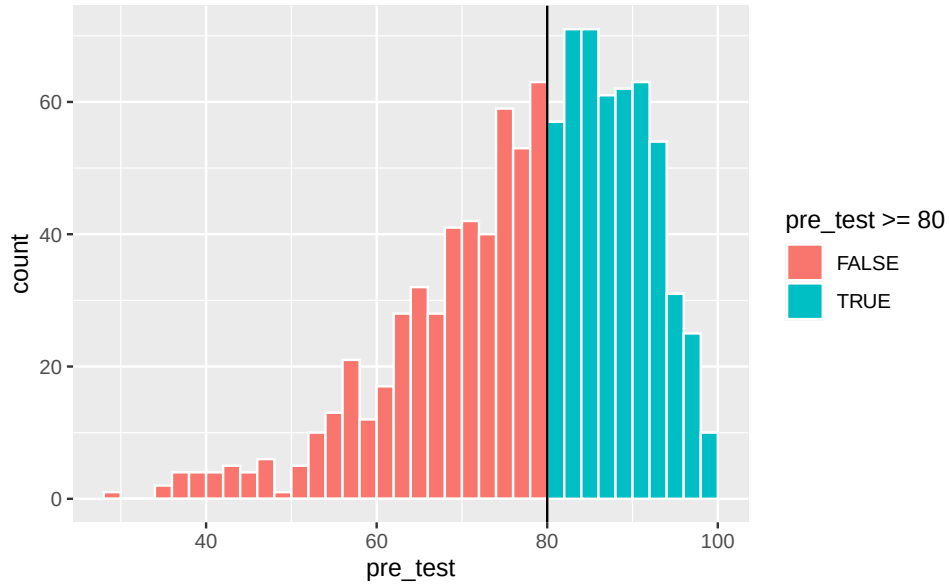
↓ 1: What's going on here? ↓

```
ggplot(program, aes(x = pre_test, y = honors, color = honors)) +
  geom_point(size = 1, alpha = 0.5,
             position = position_jitter(width = 0, height = 0.25)) +
  geom_vline(xintercept = 80)
```



↓ 2: What's going on here? ↓

```
ggplot(program, aes(x = pre_test, fill = pre_test >= 80)) +
  geom_histogram(binwidth = 2, color = "white", boundary = 80) +
  geom_vline(xintercept = 80)
```

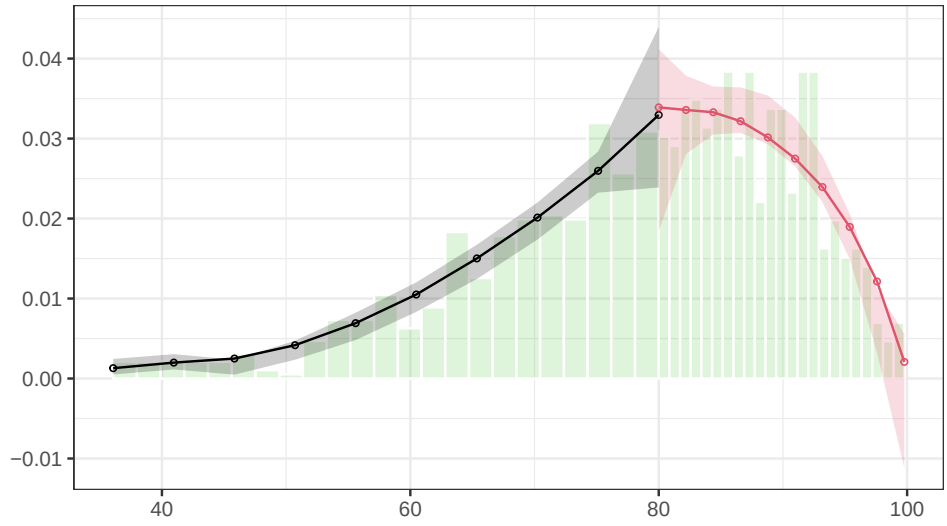



```
density_check <- rddensity(program$pre_test, c = 80)
summary(density_check)
```

```
##
```

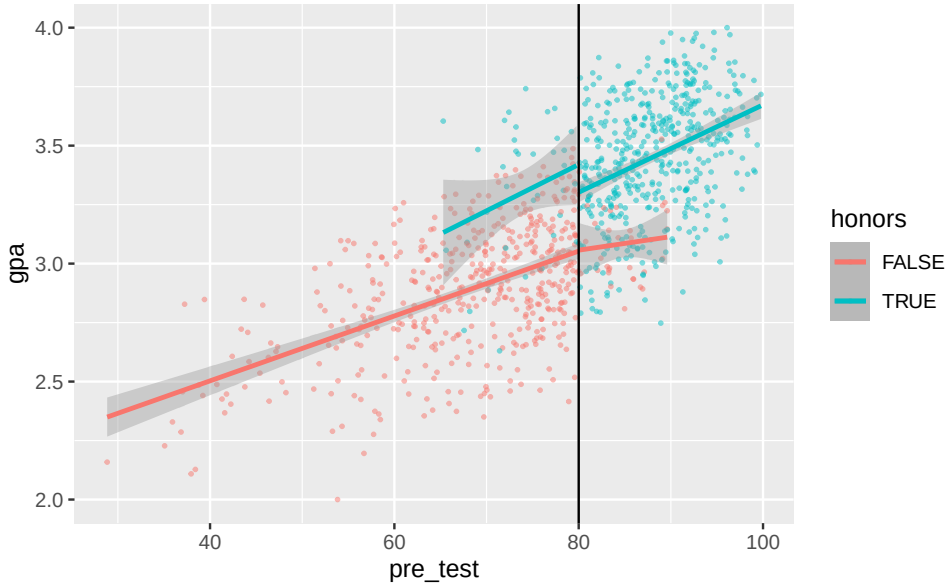
```
## Manipulation testing using local polynomial density estimation.
##
## Number of obs =       1000
## Model =               unrestricted
## Kernel =              triangular
## BW method =           estimated
## VCE method =          jackknife
##
## c = 80                Left of c           Right of c
## Number of obs         495                 505
## Eff. Number of obs    334                 418
## Order est. (p)        2                   2
## Order bias (q)        3                   3
## BW est. (h)           14.642              12.792
##
## Method                T                   P > |T|
## Robust                -0.5146             0.6068
##
##
## P-values of binomial tests (H0: p=0.5).
##
## Window Length / 2          <c     >=c    P>|T|
## 0.404                       9      11    0.8238
## 0.807                      27      21    0.4709
## 1.211                      37      37    1.0000
## 1.614                      50      50    1.0000
## 2.018                      65      57    0.5264
## 2.421                      73      76    0.8699
## 2.825                      87      85    0.9393
## 3.228                      99      98    1.0000
## 3.632                     107     116    0.5923
## 4.035                     118     129    0.5247
```

```
plot_mccrary <- rdplotdensity(rdd = density_check,
                              X = program$pre_test,
                              type = "both")
```



↓ 3: What's going on here? ↓

```
ggplot(program, aes(x = pre_test, y = gpa, color = honors)) +
  geom_point(size = 0.5, alpha = 0.5) +
  geom_smooth(data = filter(program, pre_test >= 80), method = "lm") +
  geom_smooth(data = filter(program, pre_test < 80), method = "lm") +
  geom_vline(xintercept = 80)
```



↓ 4: What's going on here? ↓

```
program <- program %>%
  mutate(above_cutoff = pre_test >= 80)

# Parametric stuff
model_parametric_5 <- iv_robust(
  gpa ~ pre_test_centered + honors | pre_test_centered + above_cutoff,
  data = filter(program,
                pre_test_centered >= -5,
                pre_test_centered <= 5))
tidy(model_parametric_5)
```

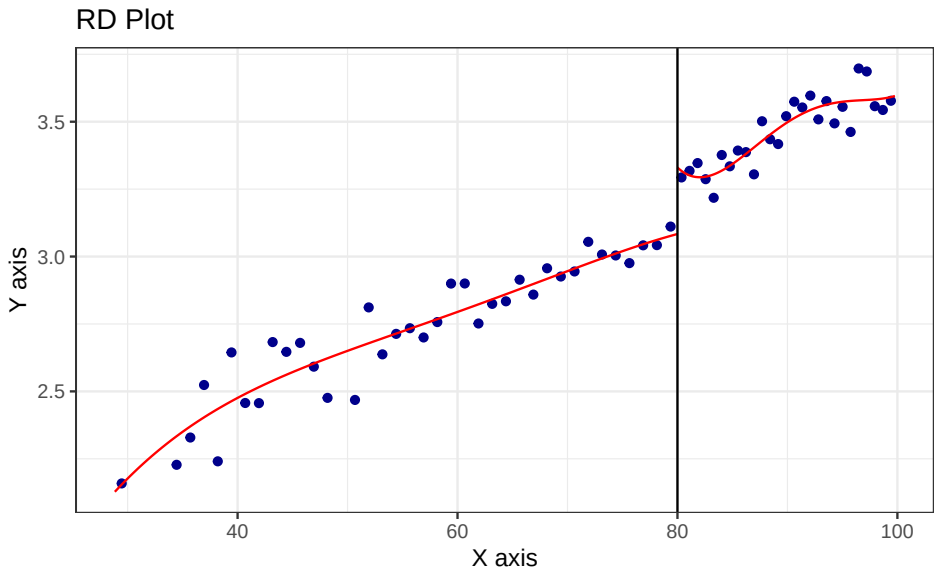
term	estimate	std.error	statistic	p.value	conf.low	conf.high	df	outcome
(Intercept)	3.061	0.036	85.62	0.000	2.991	3.132	303	gpa
pre_test_centered	0.019	0.009	2.11	0.036	0.001	0.036	303	gpa
honorsTRUE	0.227	0.069	3.30	0.001	0.091	0.362	303	gpa

```
model_parametric_10 <- iv_robust(
  gpa ~ pre_test_centered + honors | pre_test_centered + above_cutoff,
  data = filter(program,
                pre_test_centered >= -10,
                pre_test_centered <= 10))
tidy(model_parametric_10)
```

term	estimate	std.error	statistic	p.value	conf.low	conf.high	df	outcome
(Intercept)	3.076	0.027	115.40	0	3.024	3.128	576	gpa
pre_test_centered	0.017	0.004	4.71	0	0.010	0.024	576	gpa
honorsTRUE	0.224	0.050	4.50	0	0.126	0.321	576	gpa

↓ 5: What's going on here? ↓

```
# Nonparametric stuff
rdplot(y = program$gpa, x = program$pre_test, c = 80)
```



↓ 6: What's going on here? ↓

```
rdrobust(y = program$gpa, x = program$pre_test, fuzzy = program$honors, c = 80) %>%
  summary()
```

```
## Call: rdrobust
##
## Number of Obs.                 1000
## BW type                       mserd
## Kernel                   Triangular
## VCE method                       NN
##
## Number of Obs.                 495         505
## Eff. Number of Obs.            214         258
## Order est. (p)                   1           1
## Order bias  (q)                  2           2
## BW est. (h)                  7.973       7.973
## BW bias (b)                 12.132      12.132
## rho (h/b)                    0.657       0.657
## Unique Obs.                    495         505
##
## ===============================================================================
##         Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## ===============================================================================
##   Conventional     0.237     0.062     3.857     0.000     [0.117 , 0.358]
##         Robust         -         -     3.257     0.001     [0.096 , 0.388]
## ===============================================================================
```

↓ 7: What's going on here? ↓

```
rdrobust(y = program$gpa, x = program$pre_test,
         fuzzy = program$honors, c = 80, h = 7.973 * 2) %>%
  summary()
```

```
## Call: rdrobust
##
## Number of Obs.                 1000
## BW type                       Manual
## Kernel                   Triangular
## VCE method                       NN
##
## Number of Obs.                 495         505
## Eff. Number of Obs.            358         470
## Order est. (p)                   1           1
## Order bias  (q)                  2           2
## BW est. (h)                 15.946      15.946
## BW bias (b)                 15.946      15.946
## rho (h/b)                    1.000       1.000
## Unique Obs.                    495         505
##
## ===============================================================================
##         Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## ===============================================================================
##   Conventional     0.225     0.045     4.981     0.000     [0.136 , 0.313]
##         Robust         -         -     3.564     0.000     [0.109 , 0.375]
## ===============================================================================
```

```
rdrobust(y = program$gpa, x = program$pre_test,
         fuzzy = program$honors, c = 80, h = 7.973 / 2) %>%
  summary()
```

```
## Call: rdrobust
##
## Number of Obs.                 1000
## BW type                       Manual
## Kernel                   Triangular
## VCE method                       NN
##
## Number of Obs.                 495         505
## Eff. Number of Obs.            115         127
## Order est. (p)                   1           1
## Order bias  (q)                  2           2
## BW est. (h)                  3.986       3.986
## BW bias (b)                  3.986       3.986
## rho (h/b)                    1.000       1.000
## Unique Obs.                    495         505
##
## ===============================================================================
##         Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## ===============================================================================
##   Conventional     0.224     0.084     2.660     0.008     [0.059 , 0.389]
##         Robust         -         -     1.021     0.307    [-0.124 , 0.392]
## ===============================================================================
```

```r
rdrobust(y = program$gpa, x = program$pre_test,
         fuzzy = program$honors, c = 80, kernel = "epanechnikov") %>%
  summary()
```

```
## Call: rdrobust
##
## Number of Obs.                 1000
## BW type                       mserd
## Kernel                   Epanechnikov
## VCE method                       NN
##
## Number of Obs.                 495         505
## Eff. Number of Obs.            190         221
## Order est. (p)                   1           1
## Order bias  (q)                  2           2
## BW est. (h)                  6.782       6.782
## BW bias (b)                 10.623      10.623
## rho (h/b)                    0.638       0.638
## Unique Obs.                    495         505
##
## ===============================================================================
##         Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## ===============================================================================
##   Conventional     0.229     0.064     3.556     0.000     [0.103 , 0.356]
##         Robust         -         -     2.927     0.003     [0.075 , 0.381]
## ===============================================================================
```