

Econometrics 2 PS4: Regression Discontinuity

Carlos T. Estrada Arzamendi

March 21, 2024

Problem 1: (Sharp) Regression Discontinuity

“Take any dataset with covariate X and outcome Y that are related in some way. For instance, you can use the data on birth weight and smoking from here: <http://www.stata.com/texts/eacsap/>, or any other relevant dataset. Alternatively, feel free to simulate your own data. In any case, please provide an explanation of your dataset. Construct a placebo treatment by applying a rule such that $D_i = 1$ when $X_i \geq x_0$ for some x_0 . That is, modify the outcome variable Y_i for those units with $X_i \geq x_0$ by adding a constant treatment effect, for example, add one standard deviation of the outcome plus some noise (with mean zero). Include an explanation of what you ended up doing”

```
# Data Simulation
library(truncnorm)
n = 10000

data = data.table(
  gpa = rtruncnorm(n, a = 0, b = 4, mean = 3, sd = 1),
  fam_income = runif(n, min = 20000, max = 150000)
)
gpa_cutoff = 2

# if gpa is below cutoff, school provides tutoring that increases score
data[, treat := ifelse(gpa < gpa_cutoff, 1, 0)]
data[, outcome_score := gpa*300 + fam_income/1000 + treat*200 + rnorm(n, mean = 0, sd = 50)]

datasummary_skim(data, out = "markdown", title = "Summary Statistics", histogram = F)
```

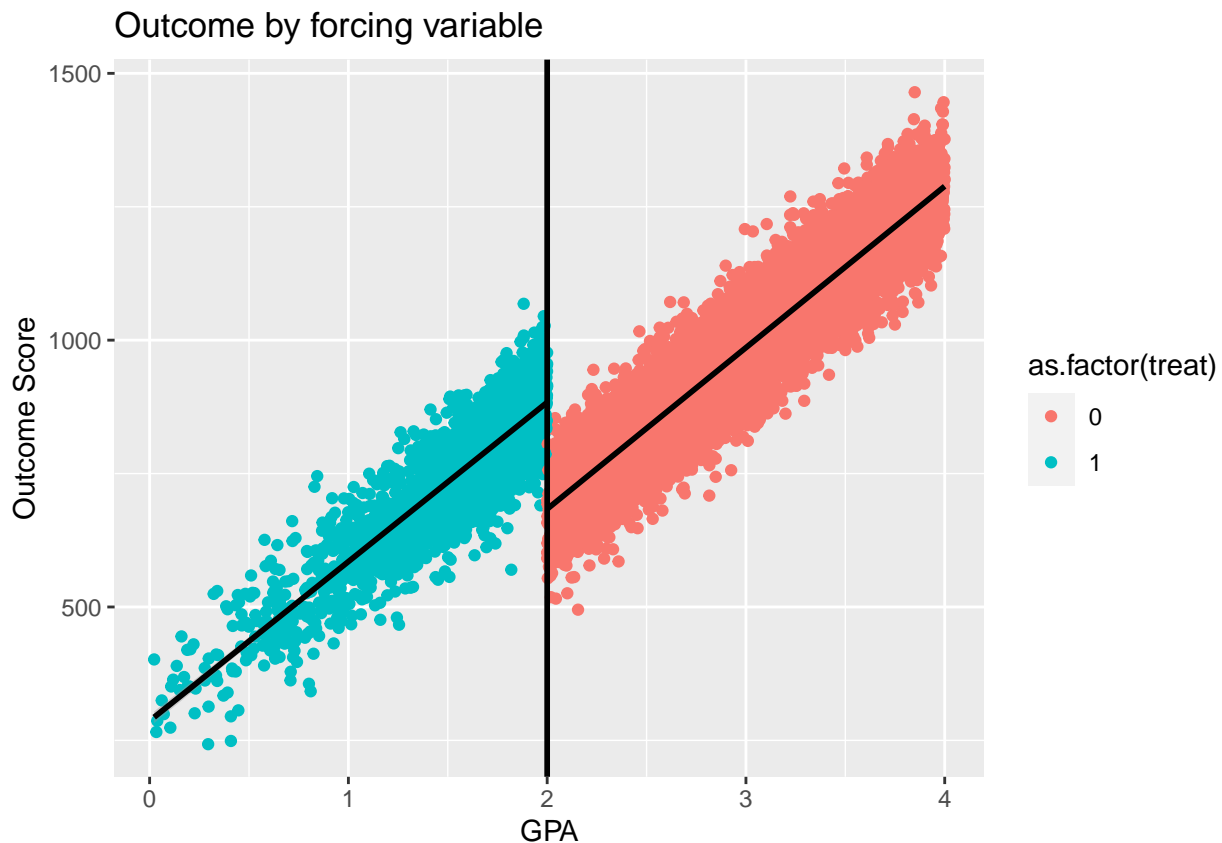
Table 1: Summary Statistics

	Unique	Missing Pct.	Mean	SD	Min	Median	Max
gpa	10000	0	2.7	0.8	0.0	2.8	4.0
fam_income	10000	0	85409.1	37744.9	20001.0	85357.9	149973.8
treat	2	0	0.2	0.4	0.0	0.0	1.0
outcome_score	10000	0	938.5	195.2	242.9	931.0	1464.6

1.1) Plot the outcome by forcing variable (the standard graph showing the discontinuity)

```
# Generate input data for output plot
```

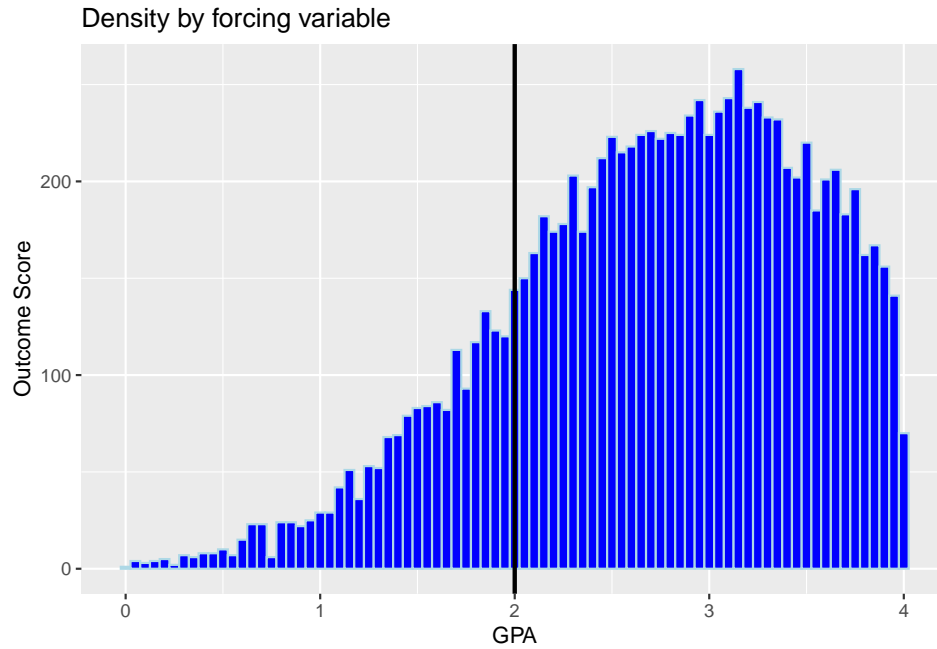
```
plot1 <- ggplot(data, aes(x = gpa, y = outcome_score, color = as.factor(treat), group = as.factor(treat))) +  
  geom_point() +  
  #geom_smooth(aes(fill = as.factor(treat))) +  
  geom_smooth(method = "lm", color = "black", formula = y~x) +  
  labs(x = "GPA", y = "Outcome Score") + ggtitle("Outcome by forcing variable") +  
  geom_vline(xintercept = gpa_cutoff, linewidth = 1)  
plot1
```



1.2) Plot the density of the forcing variable

You can also embed plots, for example:

```
plot2 <- ggplot(data, aes(gpa)) +  
  geom_histogram(fill = "blue", color = "lightblue", binwidth = 0.05) +  
  labs(x = "GPA", y = "Outcome Score") + ggtitle("Density by forcing variable") +  
  geom_vline(xintercept = gpa_cutoff, linewidth = 1)  
plot2
```



1.3) Estimate the effect using a local linear regression

```
reg3 = lm(outcome_score ~ treat + I(gpa-gpa_cutoff) + I(treat*(gpa-gpa_cutoff)) , data = data)
stargazer(reg3, type = "latex", title = "Local Linear Regression", table.placement = "H",
  header=FALSE, no.space = T, omit.stat = "f")
```

Table 2: Local Linear Regression

	<i>Dependent variable:</i>
	outcome_score
treat	199.937*** (2.751)
I(gpa - gpa_cutoff)	302.432*** (1.293)
I(treat *(gpa - gpa_cutoff))	-4.011 (3.788)
Constant	683.559*** (1.466)
Observations	10,000
R ²	0.896
Adjusted R ²	0.896
Residual Std. Error	62.952 (df = 9996)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

1.4) Estimate the effect using a local polynomial (of order 2 and 3) regression

```
reg41 = lm(outcome_score ~ treat + I(gpa-gpa_cutoff) + I(treat*(gpa-gpa_cutoff))+
          I((gpa-gpa_cutoff)^2) + I(treat*(gpa-gpa_cutoff)^2), data = data)
#stargazer(reg41, type = "latex", title = "Estimate Effect Using Local Polynomial of Order 2", table.placement = "H", header=FALSE, no.space = T, omit.stat = "f")

reg42 = lm(outcome_score ~ treat + I(gpa-gpa_cutoff) + I(treat*(gpa-gpa_cutoff))+
          I((gpa-gpa_cutoff)^2) + I(treat*(gpa-gpa_cutoff)^2)+
          I((gpa-gpa_cutoff)^3) + I(treat*(gpa-gpa_cutoff)^3), data = data)
stargazer(reg3, reg41, reg42, type = "latex", title = "Estimate Effect Using Local Polynomial of Order 2 and 3", table.placement = "H", header=FALSE, no.space = T, omit.stat = "f")
```

Table 3: Estimate Effect Using Local Polynomial of Order 2 and 3

	<i>Dependent variable:</i>		
	outcome_score		
	(1)	(2)	(3)
treat	199.937*** (2.751)	207.002*** (3.895)	205.134*** (5.056)
I(gpa - gpa_cutoff)	302.432*** (1.293)	302.894*** (5.106)	286.485*** (12.927)
I(treat *(gpa - gpa_cutoff))	-4.011 (3.788)	27.907** (11.809)	53.295** (26.452)
I((gpa - gpa_cutoff)^2)		-0.232 (2.477)	19.984 (14.840)
I(treat *(gpa - gpa_cutoff)^2)		22.565*** (7.352)	16.930 (37.061)
I((gpa - gpa_cutoff)^3)			-6.742 (4.880)
I(treat *(gpa - gpa_cutoff)^3)			12.653 (14.334)
Constant	683.559*** (1.466)	683.396*** (2.276)	686.328*** (3.112)
Observations	10,000	10,000	10,000
R ²	0.896	0.896	0.896
Adjusted R ²	0.896	0.896	0.896
Residual Std. Error	62.952 (df = 9996)	62.926 (df = 9994)	62.925 (df = 9992)

Note:

*p<0.1; **p<0.05; ***p<0.01