Econometrics 2 PS4: Regression Discontinuity

Carlos T. Estrada Arzamendi

March 23, 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 Di = 1 when Xi >= x0 for some x0. That is, modify the outcome variable Yi for those units with Xi >= x0 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"

```r
# Change to test how RMarkdown works with VSCode
# Change 2--- synced?

# Data Simulation
library(truncnorm)
```

```
## Warning: package 'truncnorm' was built under R version 4.3.3
```

```r
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)
```

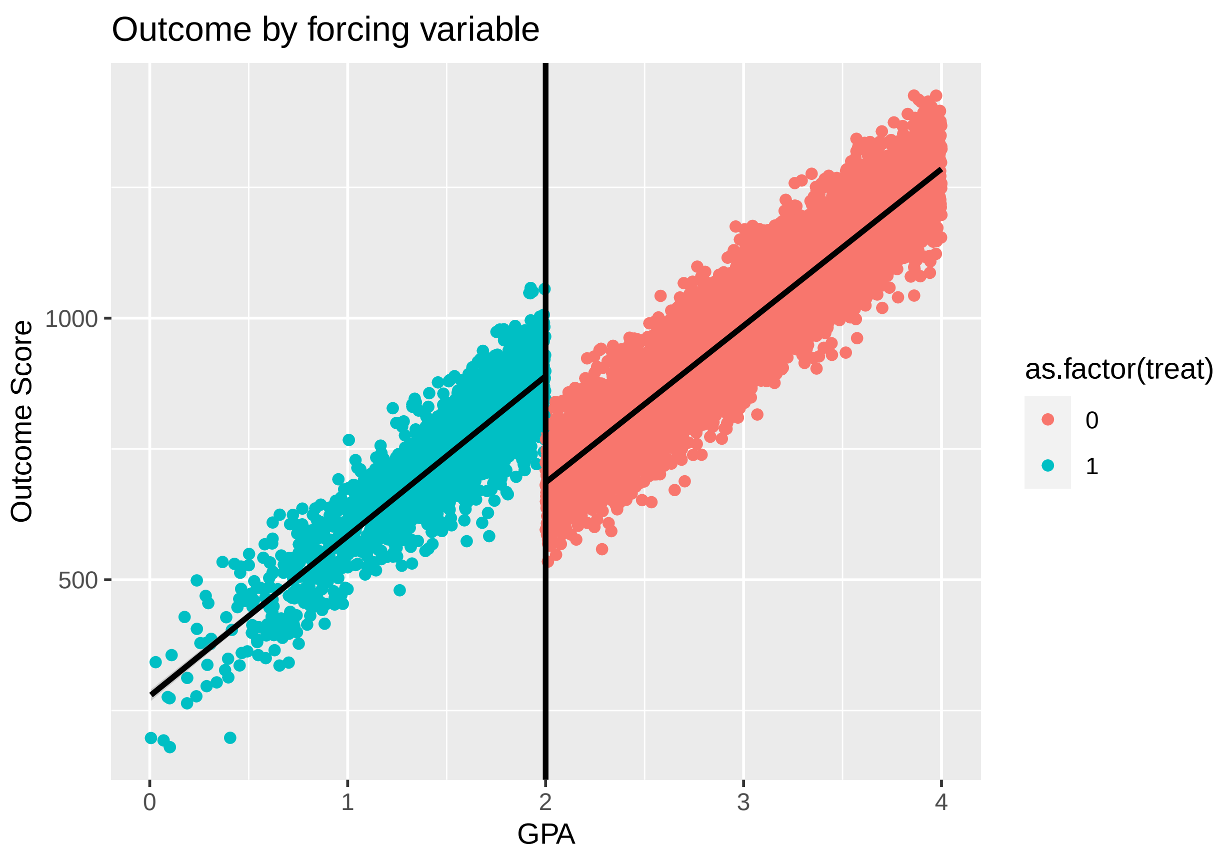
Table 1: Summary Statistics

	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	85245.7	37431.9	20018.7	85403.9	149967.3
treat	2	0	0.2	0.4	0.0	0.0	1.0

	Unique	Missing Pct.	Mean	SD	Min	Median	Max
outcome_score	10000	0	940.1	194.7	180.1	931.7	1425.4

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

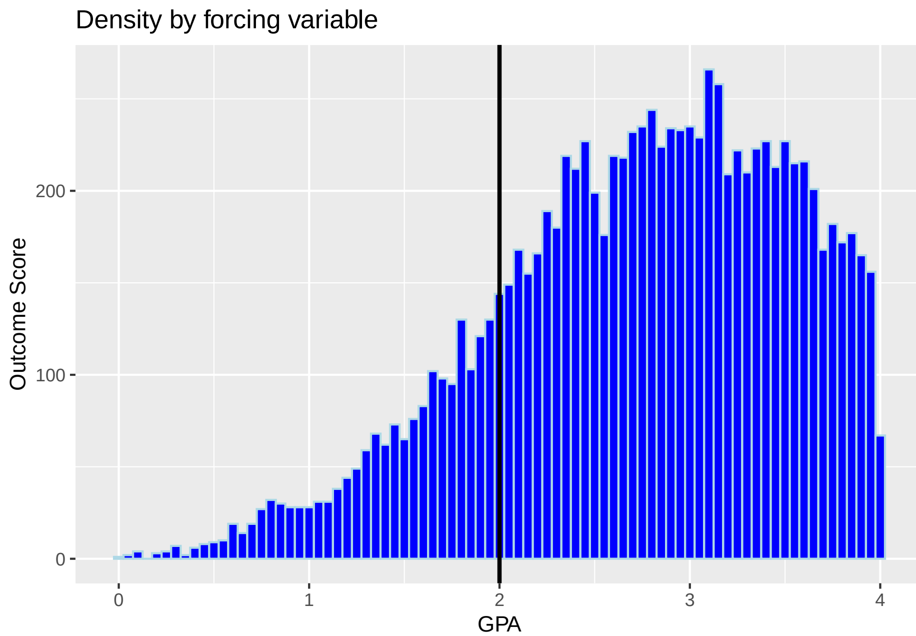
```
# Generate input data for output plot

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:

```r
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

```r
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

	Dependent variable:
	outcome_score
treat	203.537***
	(2.780)
I(gpa - gpa_cutoff)	299.659***
	(1.281)
I(treat *(gpa - gpa_cutoff))	6.113
	(3.812)
Constant	685.806***
	(1.459)
Observations	10,000
R^2	0.896
Adjusted R^2	0.896
Residual Std. Error	62.791 (df = 9996)
Note:	*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.pl
```

```
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 
          table.placement = "H", header=FALSE, no.space = T, omit.stat = "f")
```

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

	Dependent variable:		
	outcome_score		
	(1)	(2)	(3)
treat	203.537***	203.802***	206.974***
	(2.780)	(3.941)	(5.085)
I(gpa - gpa_cutoff)	299.659***	298.048***	290.306***
	(1.281)	(5.077)	(12.789)
I(treat ∗(gpa - gpa_cutoff))	6.113	11.720	58.001**
	(3.812)	(12.205)	(26.691)
I((gpa - gpa_cutoff)ˆ2)		0.805	10.347
		(2.457)	(14.674)
I(treat ∗(gpa - gpa_cutoff)ˆ2)		2.012	55.470
		(7.800)	(37.519)
I((gpa - gpa_cutoff)ˆ3)			−3.181
			(4.823)
I(treat ∗(gpa - gpa_cutoff)ˆ3)			29.224**
			(14.753)
Constant	685.806***	686.375***	687.755***
	(1.459)	(2.268)	(3.086)
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	62.791 (df = 9996)	62.797 (df = 9994)	62.791 (df = 9992)
Note:			*p<0.1; **p<0.05; ***p<0.01