Regression Discontinuity

simulation code

```
set.seed(1234)
n = 100
x = runif(n)
D = 1*(x>=0.5)

y = pnorm(x,0.5,0.1)+rnorm(n,sd=0.1)+D
cutoff = 0.5
xtilde = x-cutoff
rd = lm(y~D+xtilde)
coef(rd)
```

```
## (Intercept)           D      xtilde
##   0.2587526   1.4629172   0.7543980
```

```
summary(rd)
```

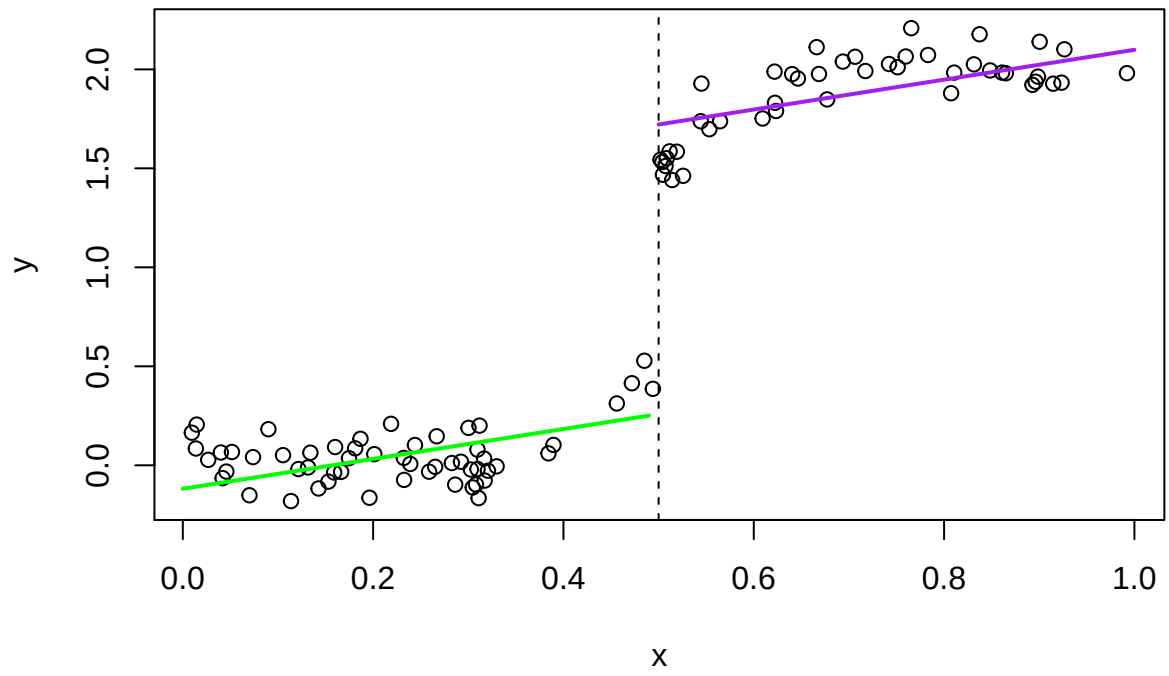
```
##
## Call:
## lm(formula = y ~ D + xtilde)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.291540 -0.107167 -0.008672  0.112488  0.313228
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.25875    0.03552   7.284 8.61e-11 ***
## D            1.46292    0.05898  24.802  < 2e-16 ***
## xtilde       0.75440    0.10582   7.129 1.81e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1442 on 97 degrees of freedom
## Multiple R-squared:  0.9765, Adjusted R-squared:  0.976
## F-statistic:  2018 on 2 and 97 DF,  p-value: < 2.2e-16
```

```
confint(rd, 'D', level=0.95)
```

```
##      2.5 %   97.5 %
## D 1.345852 1.579982
```

```
plot(x,y)
abline(v=cutoff,lty=2)
x2 = seq(0,cutoff-0.01,0.01)
y2 = predict(rd, newdata = data.frame(xtilde=x2-cutoff,D=1 * (x2>=cutoff)))
points(x2, y2, type='l', col='green',lwd=2)

x3=seq(cutoff,1,0.01)
y3=predict(rd,newdata=data.frame(xtilde=x3-cutoff,D=1*(x3>=cutoff)))
points(x3, y3, type= 'l', col='purple',lwd=2)
```



cutoff = 0.5 E[Y0|X] when D =0, E[Y0|X] when D = 1 95 percent confidence interval: 1.345852 1.579982 true mean is not equal to 0. It is very significant.

linear RD model to estimate the causal e¤ect of legal access to alcohol on death rates.

```
library(magrittr)
library(dplyr)
```

```
## Registered S3 methods overwritten by 'tibble':
##   method     from
##   format.tbl pillar
##   print.tbl  pillar

##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
mlda = read.csv('mlda.csv')
mlda = mlda %>% mutate(age = agecell - 21,
                       over21 = 1 * (agecell >= 21))
linear = lm(all ~ over21 + age + age:over21, mlda)
summary(linear)
```

```
##
## Call:
## lm(formula = all ~ over21 + age + age:over21, data = mlda)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.368 -1.787  0.117  1.108  5.341
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  93.6184     0.9325 100.399  < 2e-16 ***
## over21        7.6627     1.3187   5.811  6.4e-07 ***
## age           0.8270     0.8189   1.010  0.31809
## over21:age   -3.6034     1.1581  -3.111  0.00327 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.283 on 44 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.6677, Adjusted R-squared:  0.645
## F-statistic: 29.47 on 3 and 44 DF,  p-value: 1.325e-10
```

quadratic rather than linear specification.

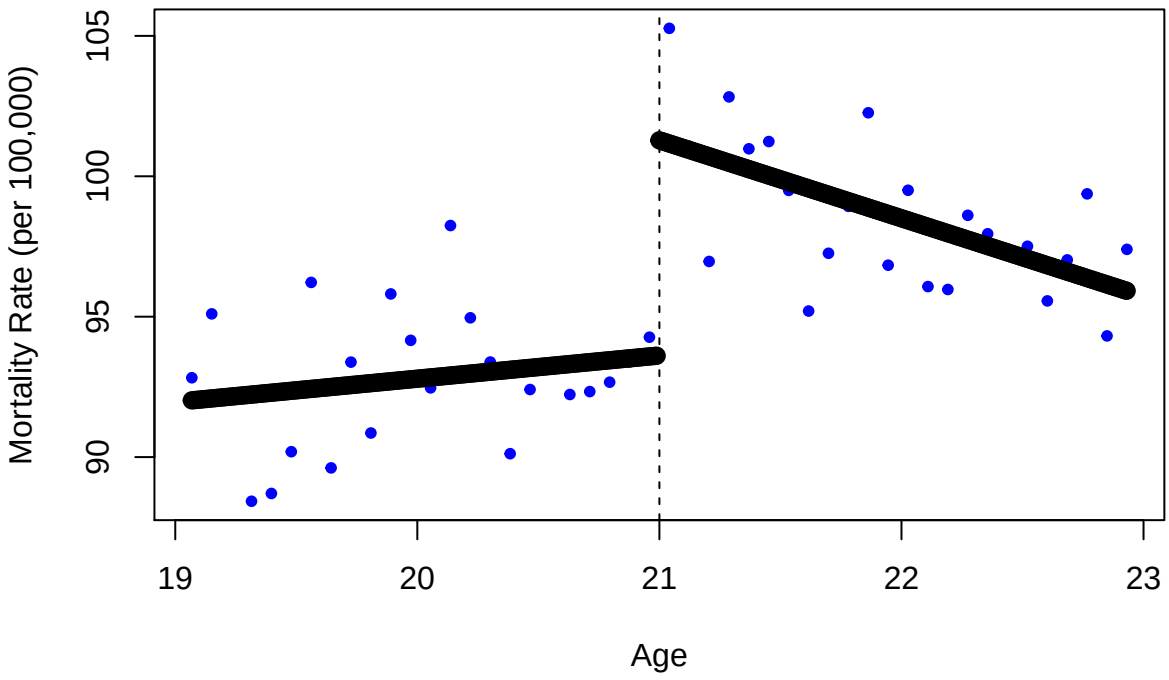
```
quadratic = lm(all ~ over21 + age + I(age^2) +
                 age:over21 + I(age^2):over21, mlda)
summary(quadratic)
```

```
##
## Call:
## lm(formula = all ~ over21 + age + I(age^2) + age:over21 + I(age^2):over21,
##     data = mlda)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.3343 -1.3946  0.1849  1.2848  5.0817
##
## Coefficients:
```

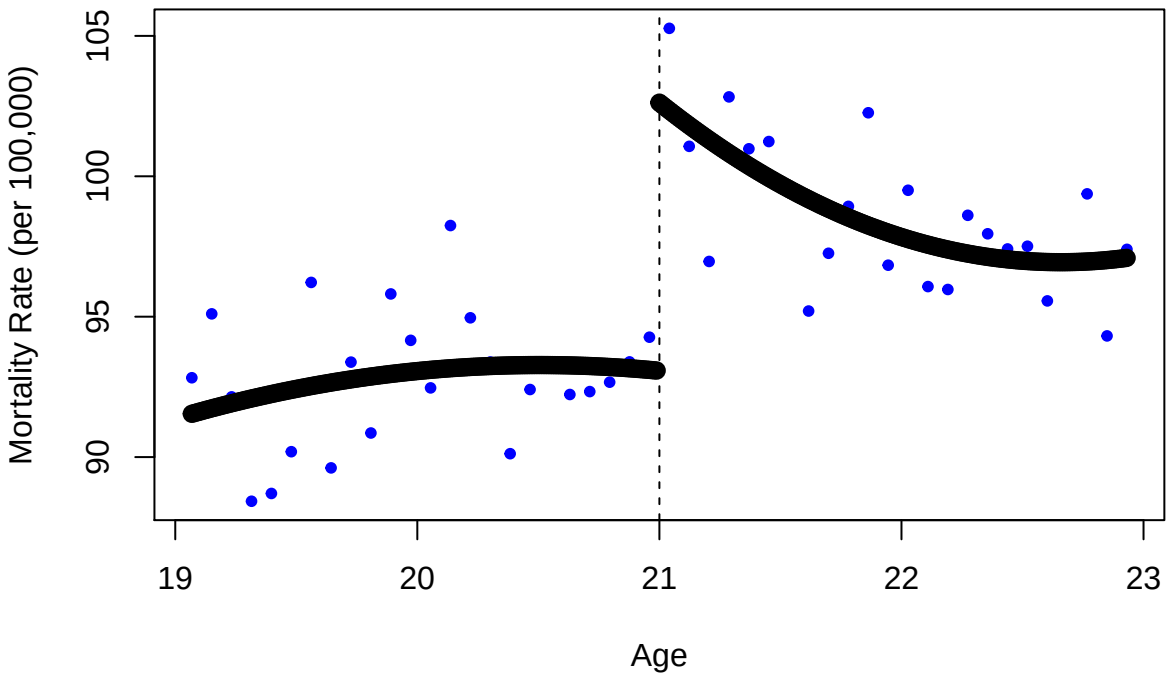
```
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      93.0729     1.4038  66.301  < 2e-16 ***
## over21            9.5478     1.9853   4.809 1.97e-05 ***
## age              -0.8306     3.2901  -0.252    0.802
## I(age^2)         -0.8403     1.6153  -0.520    0.606
## over21:age       -6.0170     4.6529  -1.293    0.203
## over21:I(age^2)   2.9042     2.2843   1.271    0.211
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.285 on 42 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.6821, Adjusted R-squared:  0.6442
## F-statistic: 18.02 on 5 and 42 DF,  p-value: 1.624e-09
```

```
make_RD_plot = function(reg, dat, inc = 0.01){
  plot(all ~ agecell, dat, xlab = 'Age',
       ylab = "Mortality Rate (per 100,000)",
       pch = 20, col = 'blue')
  abline(v = 21, lty = 2)
  x_min = min(dat$agecell)
  x_max = max(dat$agecell)
  x2 = seq(x_min, 21 - inc, inc)
  x3 = seq(21, x_max, inc)
  y2 = predict(reg, data.frame(age = x2 - 21,
                               over21 = 1 * (x2 >= 21)))
  y3 = predict(reg, data.frame(age = x3 - 21,
                               over21 = 1 * (x3 >= 21)))
  points(x2, y2, lwd = 2)
  points(x3, y3, lwd = 2)
}
```

```
make_RD_plot(linear, mlda)
```




```
make_RD_plot(quadratic, mlda)
```



RD analysis by restricting your sample to ages between 20 and 22:

```
mlda2 = read.csv('mlda2.csv')
linear2 = lm(all ~ agecell , mlda2)
summary(linear2)
```

```
##
## Call:
## lm(formula = all ~ agecell, data = mlda2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.0443 -2.0938 -0.7953  1.7555  9.1335
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  65.1067    13.7982   4.718 3.97e-05 ***
## agecell       1.4746     0.6415   2.299   0.0278 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.286 on 34 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.1345, Adjusted R-squared:  0.1091
## F-statistic: 5.285 on 1 and 34 DF,  p-value: 0.02779
```

The age between 20-22 is less significant than only the age over 21. And the R-squared is lower than before.

```
quadratic2 = lm(all ~ agecell + I(agecell^2), mlda2)
summary(quadratic2)
```

```
##
## Call:
## lm(formula = all ~ agecell + I(agecell^2), data = mlda2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9355 -2.2183 -0.1095  1.5780  7.9750
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -953.2820   351.7477  -2.710  0.01059 *
## agecell        96.3885    32.7673   2.942  0.00593 **
## I(agecell^2)   -2.2080     0.7622  -2.897  0.00664 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.978 on 33 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:   0.31,  Adjusted R-squared:  0.2682
## F-statistic: 7.413 on 2 and 33 DF,  p-value: 0.002192
```