We ultimately want to estimate the effect of D_i on Y. Suppose that D_i is determined, in part, by whether $X_i \geq c$. We call X_i the *running* or *forcing* variable. We also assume that $Y_i(0)$ and $Y_i(1)$ are related to X_i *continuously*. There shouldn't be a large and discontinuous jump in Y_i as X_i changes because we assume a smooth relationship. The probability of treatment goes up or down with X_i as it is related to c — a causal effect on whether i gets treated. If we see that Y jumps disconinously, then we can estimate or identify the effect arount the threshold value c.

For example, test score cutoffs have been used as the threshold value for regression discontinuity design. If a person failed a test, they would have to go to summer school; individuals around the cutoff served as the variation to help identify the efficacy or effect of summer school.

We will never observe multiple treatments for the same individual. The $Y_i(1)$ and $Y_i(0)$ are smooth functions of X_i. Formally, $\mathbb{E}[Y_i(0)|X_i = x]$ and $\mathbb{E}[Y_i(1)|X_i = x]$ are continuous in X. Then

$$\tau_{srd} = \lim_{X \downarrow c} \mathbb{E}[Y_i|X_i = x] - \lim_{X \uparrow c} \mathbb{E}[Y_i|X_i = x] \tag{1}$$

Rarely will we observe an individual with $X_i = c$; but in that case, we have assumed that the treatment is granted, here. We can't see what happens either $\mathbb{E}[Y_i(0)|X_i = c]$ or $\mathbb{E}[Y_i(1)|X_i = x]$ in order to estimate the true treatment effect, defined by

$$\tau_{srd} = \mathbb{E}[Y_i(1) - Y_i(0)|X_i = c] = \mathbb{E}[Y_i(1)|X_i = c] - \mathbb{E}[Y_i(0)|X_i = c]$$

We are trying to estimate Equation 1 empirically by only looking at those individuals "near" the threshold c. We can look at this in code with the generating process

$$Y_i = 1 + X_i + 2 \cdot D_i + \epsilon \quad \text{with} \quad \epsilon \sim N(0, 1/2)$$

where $i \in \{1, 2, \ldots, N\}$. In particular, we set $\tau_{srd} = 2$. Can we estimate this treatment effect using regression discontinuity? First consider the sharp RD design — everyone with $X_i > c$ gets treated and everyone else does not.

```
c <- 0.5; N <- 10000
X <- runif(N)
D <- ifelse(X > c, 1, 0)
Y <- 1 + X + 2*D + rnorm(N, sd = 0.5)
```

Now define a bandwidth b, where we restrict our attention to observations with $X_i \in (c - b, c + b)$. The following code collects the indices for these individuals.

```
b <- 0.05
lower <- X < c & X > c - b
upper <- X > c & X < c + b
```

The total number of individuals in this group should be about 10% of the total sample, or about 1,000 when $N = 10,000$, given that the X_i's are drawn from a uniform distribution. This is shown to be true:

```
length(c(which(upper), which(lower)))
```

```
[1] 1002
```

It follows from Equation (1) that we can simply difference the outcome variables for the upper and lower groups. Indeed the outcome reflects this shift – which can also be plotted.

```
mean(Y[upper]) - mean(Y[lower])
```

```
[1] 2.056059
```

Now consider the fuzzy RD design, where D_i is no longer determined *only* by X_i. Mathematically,

$$0 < \lim_{X \downarrow c} \mathbb{P}[D_i = 1 | X_i = x] - \lim_{X \uparrow c} \mathbb{P}[D_i = 1 | X_i = x] < 1 \tag{2}$$

The estimate for the treatment effect in a fuzzy regression discontinuity design is therefore given as

$$\tau_{frd} = \frac{\lim\limits_{X \downarrow c} \mathbb{E}[Y_i | X_i = x] - \lim\limits_{X \uparrow c} \mathbb{E}[Y_i | X_i = x]}{\lim\limits_{X \downarrow c} \mathbb{E}[D_i | X_i = x] - \lim\limits_{X \uparrow c} \mathbb{E}[D_i | X_i = x]} \tag{3}$$

This is equivalent to the the instrumental variables estimator with instrument $Z_i = \mathbf{1}(X_i \geq c)$. There must also be a monotonicity assumption about the way that the treatment changes with X_i. If you increase the threshold, then there is a higher hurdle to get treated, so the probability of treatment should be non-increasing. We have already assumed that the probability of treatment is increasing in the running variable X. We need to assume this so there is not strange behavior around the threshold.

```
g <- 0.5; gamma <- runif(N) + ifelse(X > c, 1, 0) * rnorm(N, sd = 0.25)
D <- ifelse(gamma > g, 0, 1)
Y <- 1 + X + 2*D + rnorm(N, sd = 0.5)

b <- 0.05
lower <- X < c & X > c - b
upper <- X > c & X < c + b

mean(Y[upper]) - mean(Y[lower]) / mean(D[upper]) - mean(D[lower])
```

```
[1] -2.531811
```

What happens when you apply the sharp design to a scenario where fuzzy RD is more appropriate?