# Regression discontinuity designs, Part II

## Learning Objective

1. Learn how to set up a regression for a basic discontinuity design.
2. Learn how to set up a regression when you anticipate non-linearities

## Learning Objective 1: Basic regression discontinuity designs

Now that now we know what regression discontinuity is and how to model linear regressions, we can combine the two in order to quantify the effects of regression discontinuity tests.

The basic equation that we follow in order to do this is:

$$y = \beta_0 + \beta_1 X + \beta_2 * I(X > D) + \epsilon,$$

where X is the main variable of interest (the variable that creates the discontinuity), D is the discontinuity threshold, $\epsilon$ is the residuals or errors. In this equation, $I(X > D)$ is a function that equals 1 if $X > D$, and it equals 0 if $X$.

This means that the prediction equation for our regression if the $X$ variable is below the discontinuity is:

$$y = \beta_0 + \beta_1 X$$

.

However, if the $X$ variable is above the threshold, the prediction equation is:

$$y = (\beta_0 + \beta_2) + \beta_1 X$$

Thus, the $\beta_2$ term test whether there is a break (or discontinuity) in the relationship between the explanatory and dependent variable right at our threshold.

## Learning Objective 2: Handling non-linearities

It is important to note that this equation assumes that the equation is perfectly linear, when in many cases, real statistical relationships are not. We can account for this in several ways. If we suspect that the slope of the equation is different on either side of the threshold, we can account for that by adding a slope argument that changes the slope if X is above the threshold:

$$y = \beta_0 + \beta_1 X + \beta_2 * I(X > D) + \beta_3 \times X \times I(X > D) + \epsilon,$$

which allows for the linear relationship between $X$ and $Y$ to be different on either side of the discontinuity.

Additionally, if we suspect that the relationship isn't linear at all and is instead curvy, we can add additional arguments for $X^2$ and $X^3$:

$$y = \beta_0 + \beta_1 X + \beta_2 * I(X > D) + \beta_3 X^2 + \beta_4 X^3 + \epsilon,$$

In all of these cases, we can essentially model two separate regression equations: one below the threshold, and one above the threshold. Then, if there is any difference between the two regressions, we can quantify the effect that the treatment has by looking at the statistical significance of $\beta_2$.

For more information about regression discontinuity, check out this video for another good explanation of some of the basics (don't worry about the part about fuzzy regression discontinuity):