**REGRESSION DISCONTINUITY ACTIVITY GUIDANCE**

**Activity learning goals:**

* Understand and explain how the Regression Discontinuity (RD) method works.
* Judge situations where RD can and cannot be applied:
  + Treatment must depend on whether the assignment variable is above or below a threshold.
  + The relationship of the assignment variable to the outcome must be continuous in the absence of treatment.
* Estimate causal effects using linear and non-linear parametric RD models.

**Introducing the activity:**

The Adams Scholarship was launched in Massachusetts in 2005. It gave small awards to students who exceeded a particular district-specific test score if they attended a public 4-year college in Massachusetts. In the scatter plots shown in the worksheet (reproduced from Goodman, 2008), GAP represents the number points above (+) or below (-) the required score. The y-axis is the enrollment rate for students that have a particular GAP. The plots show how the overall, public, and private enrollment rates varied with test score before and after the program was implemented.

**Guiding students during the activity:**

1. *What explains the upward trend in the upper left figure?*

There will be at least a few groups that need help interpreting the graphs, but once they are clear, most students quickly recognize that higher test scores make admission to college more likely.

1. *Why would regressing 2005 enrollment on a dummy for receipt of the scholarship in 2005 give a poor estimate of the program’s effect?*

Nudge groups that are stuck on this question by asking “If there were no effect of the program at all, what would you expect the sign on this dummy variable to be?”

1. *What’s true about A, C, and E but isn’t true for D and F?*

We want students to notice the discontinuous jump up at the eligibility threshold in the post-treatment period, that does not exist before the program goes into effect. Some groups need to be encouraged to compare C to D and then E to F.

1. *Based on figures B, D, and F, what are the effects of the program?*

Most groups that answer Q3 correctly also recognize that the magnitude of the jump across the threshold is an estimate of the program effect. We ask groups that answer this quickly to think about whether this is an estimate of the Average Treatment Effect (ATE) or whether it is only applicable to students near the threshold. This primes them for a later discussion of the Local Average Treatment Effect (LATE).

1. *Write down a regression model that allows a linear effect of GAP and a potential discontinuous jump at the eligibility threshold (GAP=0).* *Which coefficient represents the effect of the program?*

While the first set of questions involve building intuition for RD, the second set has students explore models that allow for formal estimation. The simplest is the one we are looking for here:

epresents the probability of attending college, is 1 when the test score is below the threshold, and is 0 when it is below. When groups are struggling, we ask them what terms would capture a linear effect of GAP and a discontinuous jump at the threshold. We want students to recognize that in any of the models they write down for Q5, Q6, and Q7, the effect of the program is the coefficient on the threshold dummy variable.

1. *Note that the underlying effect of GAP on college attendance, especially at public colleges, may be nonlinear. Write down a regression model that allows for a quadratic effect of GAP and a potential discontinuous jump at the eligibility threshold (GAP=0). Which coefficient represents the effect of the program?*

We are looking for students to add a quadratic term to the specification developed above:

1. *Write down a regression model that allows for a linear effect of GAP, a potential discontinuous jump at the eligibility threshold (where GAP=0),* ***and*** *allows the slope to be different on each side of the threshold.* *Which coefficient represents the effect of the program?*

To answer this question, students must combine what they’ve learned so far about RD with what they’ve learned about interaction terms. Specifically, they must recognize that an interaction can be used to let the effect of GAP differ for students above and below the threshold. Some students simply add the interaction:

We ask students what the slope of the regression line is on each side of the threshold ( and in this case) and make sure they recognize that the effect of the treatment is still the coefficient on the threshold dummy variable. Some students write down an equivalent model that is somewhat easier to interpret:

Here the slope to the left of the threshold is and the slope to the right is , while the effect of the program is still the coefficient on the threshold dummy variable.

**Wrapping up the activity:**

This activity is best implemented in two stages. Start by giving them a worksheet containing the figure and the first four questions and focus on building intuition. At the end of the first stage, make it very clear that the discontinuous jump is our RD estimate of the effect of the program. You should also discuss the substance of this particular study: The Adams Scholarship induced a fair amount of switching of students from private to public colleges, but it did not result in an increase in the total number of high school graduates attending a 4-year college. This is also a good opportunity to connect the econometrics they are learning in this class to the theory they may have learned in other classes. In particular, you can point out that public and private college are substitutes goods, and when the program reduces the price of a public college education, we shouldn’t be surprised that many students who would have gone to a private school switch to the lower priced good.

After the second stage of the activity (Q5-Q7) we carefully write down correct models for each of the questions and interpret their coefficients. This is an excellent time to discuss the consequences of modeling the underlying relationship between GAP and college attendance as linear when it isn’t—In the pre-treatment period (2004) it is easy to misinterpret deviations from linearity for public college enrollment as discontinuous effects. We finish the activity with a discussion of what would be different if the threshold was not zero. Suppose is the test score and is the eligibility threshold. We need a new model in order to allow the slope to differ on each side of the threshold:

Explaining why it is necessary to subtract from is far easier once students have a solid understanding of the case where the threshold is zero.