Module 9: Glossary

rdd: Regression Discontinuity Estimation
Provides the tools to undertake estimation in Regression Discontinuity Designs. Estimation is accomplished using local linear regression. A function is also included to test the assumption of no-sorting effects. Full details are available here: https://cran.r-project.org/web/packages/rdd/rdd.pdf

Regression Discontinuity designs

- **DCdensity(runvar, cutpoint, bin = NULL, bw = NULL, verbose = FALSE, plot = TRUE, ext.out = FALSE, htest = FALSE)**

Function to implement the McCrary (2008) sorting test.
The arguments are:

- **runvar –** numerical vector holding the running variable
- **cutpoint –** the cutpoint, which defaults to 0
- **bin –** the bin width, which defaults to 2*sd(runvar)*length(runvar)^(-.5)
- **bw –** the bandwidth to use (defaults to the value given in McCrary, 2008)
- **verbose –** a logical operator, specifying whether to print diagnostic information to the console
- **plot –** logical argument wondering whether to plot the histogram and density estimation
- **ext.out –** logical flag, which if TRUE, returns extended output. If FALSE, only the p-value is returned. Defaults to TRUE
- **htest –** logical flag indicating whether to return an "htest" output

- **RDestimate(formula, data, subset = NULL, cutpoint = NULL, bw = NULL, kernel = "triangular", se.type = "HC1", cluster = NULL, verbose = FALSE, model = FALSE, frame = FALSE)**

Function to calculate the Regression Discontinuity estimate. Supports both sharp and fuzzy RDD, using the AER package for 2SLS regression under the fuzzy design. Local linear regressions are performed on either side using the Imbens-Kalyanaraman optimal bandwidth calculation.

The arguments are as follows (most are optional:
- **formula –** the formula of the RDD. This is supplied in the format of y ~ x for a simple sharp RDD, or y ~ x | c1 + c2 for a sharp RDD with two covariates. Fuzzy RDD may be specified as y ~ x + z where x is the running variable, and z is the endogenous treatment variable. Covariates are then included in the same manner as in a sharp RDD
- **data –** an optional data frame (if variables above are unspecified separately)
- **subset –** an optional vector specifying the subset of observations to be used

- **cutpoint –** the cutpoint, which defaults to 0
- **bw –** a numeric vector specifying the bandwidths at which to estimate the RD. Defaults to a calculation using the Imbens-Kalyanaraman method, and then the RD is estimated with that bandwidth, half that bandwidth, and twice that bandwidth. If only a single value is passed into the function, the RD will similarly be estimated at that bandwidth, half that bandwidth, and twice that bandwidth
- **kernel** – a string specifying the kernel to be used in the local linear fitting. "triangular" kernel is the default and is the "correct" theoretical kernel to be used for edge estimation as in RDD (Lee and Lemieux 2010). Other options are "rectangular", "epanechnikov", "quartic", "triweight", "tricube", "gaussian" and "cosine". Must be specified.
- **se.type** – this specifies the robust SE calculation method to use. Options are, as in vcovHC, "HC3", "const", "HC", "HC0", "HC1", "HC2", "HC4", "HC4m", "HC5". This option is overriden by **cluster** (see below).
- **cluster –** an optional vector specifying clusters within which the errors are assumed to be correlated. This will result in reporting cluster robust SEs. This option overrides anything specified in **se.type**. It is suggested that data with a discrete running variable be clustered by each unique value of the running variable (Lee and Card 2008).
- **verbose –** will provide some additional information printed to the terminal
- **model –** logical. If TRUE, the model object will be returned
- **frame -** logical. If TRUE, the data frame used in model fitting will be returned