Regression Discontinuity

1. We replicate some results in the paper by Angrist and Lavy: "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement", *Quarterly Journal of Economics*, 114, 533-575.

   We will use the dataset on 5th grade students, **AngristLavy5.gdt**. The file includes information on average test scores in each class (`avgmath` and `avgverb`), the spring class size (`classize`), beginning-of-the-year enrollment in the school (`enrol`), a town identifier (`towncode`), and an index of students' socioeconomic status that the authors call percent disadvantaged (`tipuach`). This study is limited to pupils in the Jewish public school system, including both secular and religious schools. The authors eliminate observations with class size greater than 44, with enrollment below 6, and with no students taking the tests (`mathsize`).

   The paper exploits the fact that class size in Israel schools is capped at 40 students (what is known as Maimonides' rule), to identify the causal effect of class size on test scores. Maimonides' rule is used to identify the effects of class size because the rule induces a discontinuity in the relationship between enrollment and class size when enrollment reaches certain values. Since this discontinuity is the source of identifying information, they define a "discontinuity sample" restricted to schools with enrollments in a range close to the points of discontinuity. This discontinuity sample is defined to include only schools with enrollments in the set of intervals [36,45], [76,85], [116,125].

   i) Restrict the sample imposing the same conditions as the authors.

   ii) Define a dummy variable for the observations in the discontinuity sample. How many classes are part of the discontinuity sample?

   iii) (Extra) Provide the following descriptive statistics for class size: mean, standard deviation, 10th, 25th, 50th, 75th, and 90th percentiles.

   In addition, report average values and standard deviations for: enrollment, percent of disadvantaged, average verbal test scores, and average match test scores for the entire sample and for the discontinuity sample. Compare the results for both groups.

   iv) In Table II the authors compare OLS estimates from three different models. Estimate the three models only for **math test scores**: a simple model with class size as the only regressor, a model controlling also for the percentage of disadvantaged students, and a model controlling both for the percentage of disadvantaged students and enrollment. Comment the results and discuss the potential biases in these regressions.

   v) Compute the predicted average class size in each school s based on the "Maimonides' Rule":

$$f_s = \frac{enrol_s}{int\left[\frac{enrol_s - 1}{40}\right] + 1}$$

   where $int(n)$ is the largest integer less than or equal to n.

vi) In Table III the authors report the reduced-form relationship between predicted class size and also test scores for several specifications. Replicate columns (2) and (6) and comment the results. Column (2) corresponds to a model with class size as the dependent variable and f_s, the percentage of disadvantaged students and enrollment as controls. Column (6) corresponds to a model with math scores as the dependent variable and f_s, the percentage of disadvantaged students and enrollment as controls.

vii) In Table IV, the authors compare 2SLS estimates from different models using f_s as an instrument for class size. Compute 2SLS estimates from a model for math test scores controlling for class size, the percentage of disadvantaged students, and enrollment (column (8) in Table IV). Interpret the results and compare the estimated effect of class size with the obtained using OLS.

viii) A key issue in the RD design is to capture the relationship between enrollment and test scores. If the relationship is non-linear the previous model may be wrong. That's why the authors include a quadratic in enrollment as an alternative specification. Estimate the model again including this variable as an additional regressor.

ix) Estimate using 2SLS the model for math test scores controlling for class size, the percentage of disadvantaged students, and enrollment for the discontinuity sample (column (12) in Table IV). Compare these results with those for the full sample.