**Randomized Controlled Trials (RCT) and Difference-in-Difference (DiD) Test**

# STAR case – RCT estimator

The data for the following exercises comes from the Project STAR: A Randomized Experiment of the Impact of Class Size Reductions on Pupil Achievement. This was a 4-year experiment in Tennessee designed to evaluate the effect of class size on learning. Each participating school had at least one control group class and one treatment group.

Use the data in star.dta to estimate the following regression models.

* 1. The dummy variable “sck” indicates whether students were in a small class. Find the mean, standard deviation, max and min values.
  2. Run a regression of the test score in kindergarten on that dummy variable that indicates whether students were in a small class.
  3. Show that the OLS estimate of the intercept in this regression will be the sample mean of the test score for those who are in the control group and that the coefficient on (small class indicator) will be the difference between the sample mean test score for those in treatment and control groups.
  4. What does this regression say about the impact of class size reductions on students’ performance? Estimate the regression, but now use robust standard errors. Why does the coefficient remain the same? Why does the standard error change?
  5. Run the following regression:

(1)

Then run the following regression:

(2) where schidkn is school identification dummy variables.

* 1. Generate a set of dummies for the school ID (), put it in the other covariates and run the regression again.
  2. What would you expect to happen to the estimate of the treatment effect for small class () in equation (2) compared to in equation (1)? Is this what happens?
  3. Consider the following equation:

(3)

where .

Now consider a hypothesis test of . Does F-test reject the null of . What does this test imply for the estimation of ?

* 1. Consider the following equation:

(4)

where

Now consider a hypothesis test of γ=0. Does F-test reject the null of .

Is assignment to a small class random?

# Worker Compensation Case

Background The data set for this exercise comes from the paper by Bruce D. Meyer, W. Kip Viscusi, David L. Durbin Workers’ compensation and injury duration: Evidence from a Natural Experiment, published in The American Economic Review, Vol. 85, No. 3, p. 322-340.

Idea MVD studied the length of time that an injured worker receives workers’ compensation. On July 15, 1980, Kentucky raised the cap on weekly earnings that were covered by workers’ compensation. An increase in the cap has no effect on the benefit for low-income workers, but this makes it less costly for a high-income worker to stay on worker’s compensation. Therefore, the control group is low-income workers, and the treatment group is high-income workers; high-income workers are defined as those who were subject to the pre-policy change cap. Using random samples both before and after the policy change, MVD were able to test whether more generous workers’ compensation causes people to stay out of work longer. They started with difference-in-difference analysis, using as dependent variable.

Use the data in injury.dta for the following questions.

* 1. Load the data and describe the data.
  2. Estimate the differences-in-differences model with regression with and without robust standard errors for Kentucky data.
  3. Re-estimate (ii) using Kentucky data by adding explanatory variables, male, married, a full set of industry and injury type dummy variables. Use robust standard errors. How does the estimate on the interaction term, (with is equivalent to ), change when these other factors are controlled for? Is the estimate still statistically significant?
  4. What can you say about the small R-squared from parts (2) and (3)?
  5. Estimate (2) using Michigan data using robust standard errors. Compare the estimates on the interaction term for Kentucky and Michigan. Does Michigan estimate statistically significant? What do you make of this?