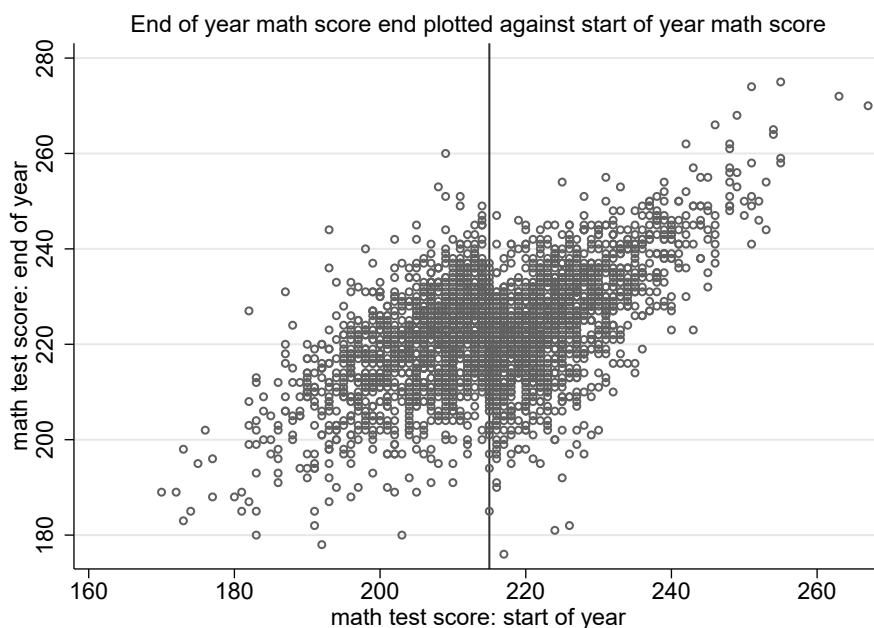


## Problem set 13: week 17

Suggested solutions. The scatter plot below plots the end of year math score against the start of year test score. If students scored 215 and below in the start of year test (the grey line) they had to attend extra tuition classes. Those who score above 215 didn't attend extra tuition classes.



- (a) We could start by carrying out a t-test: Mean end of year test score for the 1295 students who took extra classes was 218.72 with a standard deviation of 11.38. Mean end of year test score for the 1472 students who didn't take extra classes was 225.91 with a standard deviation of 12.37. The null of no mean difference in test scores is very strongly rejected. However, this does not imply the extra tuition had a (negative) causal effect on end of test score. In this situation, it is quite clear that the selection effect will be negative: those offered the extra tuition were those who scored lower at the start of the year. Therefore, one would expect the average observed effect to be below the average causal effect.
- (b) To overcome this you could estimate a regression model with an RDD specification:  $end_i = \alpha + \beta T_i + \gamma start_i + \epsilon_i$ , where  $end_i$  is the end of year score,  $T_i = 1$  if you receive the treatment and zero otherwise and  $start_i$  is the start of year score.  $\beta$  gives the treatment

effect (squared  $start_i$  terms and/or interactions between the  $T_i$  and  $start_i$  may also be added, for example).  $H_0 : \beta = 0$  versus  $H_1 : \beta \neq 0$ .  $\beta$  will pick up the causal effect under the assumption that those just above and below the cut-off are similar (the basis of RDD): That is  $E[\epsilon|T] = E[\epsilon]$ .

- (c) To check for randomisation you could see if observable characteristics balance on each side of the cut-off. Under the assumption that those just above and below the cut-off are similar (the basis of RDD) you would expect the covariates to show no statistical differences between the groups.
- (d) It's also worth looking at placebo test: i.e. there aren't just discontinuities everywhere (not just at the 215 cut-off where the treatment is).
- (e) You could also estimate more flexible regression functions as shown in the do file.