

Econometrics II. Assignment 4: Treatment evaluation and LATE

Due date: Sunday, January 30, 11.59 pm. Hand in your solutions as a **single .pdf file** including your code via Canvas. Include your R (or any other language) code by using R Markdown (preferred) or by using the package "minted" in your .tex file (see a template on Canvas). Each team has to come up with a unique name (without names or student numbers). Both teammates have to submit solutions via Canvas.

Question 1: Judges and prison sentences

A researcher is interested in the effect of prison sentence on later criminal behavior. The researcher collects data from one particular court, which has two active judges. The researcher learns that cases are assigned randomly to the judges. However, judge Jones convicts someone to a prison sentence in 70% of the cases while judge Smith only convicts someone to a prison sentence in 40% of the cases. Next, the researcher collects information on whether or not an individual is arrested in the three years after the trial.

Judge Sentences	Jones		Smith	
	Prison	Other	Prison	Other
Cases	70%	30%	40%	60%
Future arrests	40%	60%	20%	50%

- (i) Use the Wald estimator to compute the causal effect of a prison sentence on the probability of being arrested later.
- (ii) What is the interpretation of the estimated effect? And for which fraction of the population does this causal effect hold?
- (iii) Explain what an always taker is in this setting and which fraction of the population are always takers?

Question 2: Eating and exams

A teacher is interested in how important eating and drinking is for academic achievement. Therefore, the teacher decides to run a field experiment. The teacher randomizes students into two groups. The control group is allowed to have breakfast and snacks before the exam. The treatment group is only allowed to have a glass of milk in the morning and drink water during the remaining time before the exam. The key outcome variable of interest is passing the exam. In past years about 50% of the students passed the exam. The teacher thinks that not eating or drinking will reduce the passing rate with 10% points. The teacher is aiming for 70% power in the field experiment and a significance level of 5% (recall $t_{0.3} = -0.524$ and $t_{0.975} = 1.960$). The teacher randomizes students equally over the treatment and control group.

- (i) Perform a power calculation for the number of students that the teacher should include in the field experiment.

The teacher is afraid that some people in the treatment group will actually have breakfast. Furthermore he assumes that all people in the control group will have breakfast.

- (ii) The teacher assumes that 20% of the students randomized in the treatment group will actually have breakfast. How does this change the number of students required to participate in the field experiment?

Question 3: Flu shots for young children

A researcher is interested in studying the effects of providing flu shots for young children. Therefore, the researcher sets up a randomized experiment in which about 20% of the children are not entitled to a flu shot. The remaining 80% of the children are supposed to get a flu shot. The dataset `assignment4.csv(.dta)` contains the following variables:

Variable	Description
<code>treatgroup</code>	1 if a child is in the treatment group
<code>flushot</code>	1 if a child got a flu shot
<code>flu</code>	1 if a child got the flu
<code>genderchild</code>	1 if a child is a female
<code>nationality</code>	1 if a child is native
<code>agemother</code>	age of the mother at birth
<code>educmother</code>	years of education of the mother
<code>married</code>	1 if the mother is married
<code>housincome</code>	monthly household income

- (i) Compute for the children assigned to the control group the variance in flu incidence. If the researcher aims at reducing flu incidence by 0.05, how many children should participate in the randomized experiment.

Unfortunately for the researcher, there was a shortage of flu shots, and therefore not all children in the treatment group could actually get a flu shot. The variable *FluShot* indicates which children got a flu shot.

- (ii) Compute which fraction of the children in the treatment group actually received a flu shot. What is the implication for the power analysis of the experiment?

The researcher hopes that within the treatment group the flu shots are randomly assigned to children. Therefore, the researcher considers the background characteristics of the child, the mother, and the household.

- (iii) Make a table with summary statistics for (1) the control group, (2) the treated treatment group, and (3) the untreated treatment group. What do you conclude?

The researcher first focuses on only those children randomized in the treatment group. The researcher specifies the linear regression model:

$$Flu_i = \beta_0 + \beta_1 FluShot_i + u_i$$

- (iv) Estimate this model using OLS. Next, include subsequently the individual characteristics. What do you learn from these regressions?

The researcher returns to the specification:

$$Flu_i = \beta_0 + \beta_1 FluShot_i + u_i$$

but decides to instrument whether or not a child actually had a flu shot by the assignment to the treatment group.

- (v) Use 2SLS to estimate β_1 and check the robustness with respect to adding individual characteristics.

Now consider the first-stage regression:

$$FluShot_i = \gamma_0 + \gamma_1 TreatmentGroup_i + v_i$$

- (vi) Estimate the first-stage regression using OLS. Are you afraid of a weak instruments problem?
- (vii) Explain why in this case the local average treatment effect is the same as the average treatment effect on the treated.