

# Econometrics II - Assignment 3

Floris Holstege, Stanislav Avdeev

## Problem 1

Before we answer this question, let's lay out some basic terms that are useful for understanding our subsequent. We define the average treatment effect (ATE) as:

$$ATE = \mathbb{E}(\delta) = \mathbb{E}(Y_1^* - Y_0^*) = \mathbb{E}(Y_1^*) - \mathbb{E}(Y_0^*).$$

Where  $Y_1^*$  is the latent variable of interest for the group that received treatment. For this group,  $D_i = 1$ .  $Y_0^*$  is the latent variable of interest for the group that did not receive treatment. We only observe  $\mathbb{E}(Y_1^*)$ , since our control group does not receive the treatment. Given we only observe the effect when  $D_i = 1$ , we define the average treatment effect of the treated as (ATET).

$$ATET = \mathbb{E}(\delta|D = 1) = \mathbb{E}(Y_1^* - Y_0^*|D = 1) = \mathbb{E}(Y_1^*|D = 1) - \mathbb{E}(Y_0^*|D = 1).$$

If the treatment is assigned randomly, then  $(Y_{1i}^*, Y_{0i}^*) \perp D_i$ . In this case,  $ATE = ATET$ , since there is no significant difference between the characteristics of the treatment and control group. For this question, we assume that the treatment has been randomly assigned, and thus that  $ATE = ATET$ .

- Average treatment per group:  $ATE_{purple} = 9 - 7 = 2$ ,  $ATE_{blue} = 13 - 8 = 5$ ,  $ATE_{green} = 10 - 9 = 1$
- Average treatment for the full population:  $ATE = \mathbb{E}(Y_1^*) - \mathbb{E}(Y_0^*) = 10.625 - 7.875 = 2.75$
- Average treatment for the treated:  $ATE = ATET = 2.75$

## Problem 2

### I)

Before discussing descriptive statistics, we need to make several assumptions. First, we assume all variables except for total number of credits, an employment status, and an amount of efforts were measured before the randomization. Thus, we can use variables that were measured prior to the randomization to check whether group characteristics are balanced or not.

The table below suggests that there are no differences for an average number of years of education for both parents and for high-school math score between the control and two treatment groups. We can conclude that randomization was successfully done. Further, we are going to look at the following outcomes: a) percentage of students who completed all first-year courses and b) total number of credit points after one and three years of studies.

## II)

As we run a randomized experiment, the orthogonality assumption of the OLS regression is held. Thus, we are using conventional standard errors as they are more efficient. The basic model is the following:

$$COMPLETE_i = \alpha_0 + \alpha_1 HIGH_{REWARD}_i + \alpha_2 LOW_{REWARD}_i + \epsilon_i$$

We also include the following regressors to estimate an extended model:

$$COMPLETE_i = \alpha_0 + \alpha_1 HIGH_{REWARD}_i + \alpha_2 LOW_{REWARD}_i + \alpha_3 EDU_{FATHER}_i + \alpha_4 MATH_i + \alpha_5 ASSESSMENT_i + \epsilon_i$$

Columns (1) and (2) of Table 1 below shows that there is no statistically significant differences between control and two treatment groups in the probability of completing all first-year courses. The results do not change when we include control variables in the model. Based on the results we conclude that financial incentives do not affect the probability of successfully finishing the first year of the studies.

Table 1:

	<i>Dependent variable:</i>		
	pass		
	(1)	(2)	(3)
categoryHigh-reward	0.046 (0.064)	0.048 (0.058)	0.056 (0.059)
categoryLow-reward	0.007 (0.064)	0.015 (0.057)	0.023 (0.058)
math		0.119*** (0.018)	0.124*** (0.018)
fyeduc		-0.001 (0.007)	0.0003 (0.007)
p0		0.249*** (0.095)	0.170* (0.098)
effort			0.008*** (0.002)
job			-0.062 (0.060)
Constant	0.195*** (0.045)	-0.576*** (0.128)	-0.678*** (0.144)
Observations	249	245	230
Adjusted R <sup>2</sup>	-0.006	0.192	0.240

*Note:*

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

## III)

In the third model we are going to include variables that were measured after the randomization process, an employment status while studying and the amount of study effort. In line with previous results, column (3) of Table 1 yields an insignificant effect of getting financial rewards on completing all first-year courses.

However, an employment status while studying and the amount of study effort are potentially endogenous variables as they can be correlated with a treatment status of a student. One might argue that a possibility to earn 500 or 1500 guilders could have influenced a student's decision to enter the labour market while study or not. Moreover, it could have influenced an amount of efforts students dedicates to their studies, as more efforts mean higher probability of getting financial rewards. Thus, these two variables are “bad” controls that one should not include in the model.

#### IV)

We are going to use a linear probability and a logit models with only a financial dummy reward and other exogenous control variables included. We want to use a logit model to check whether there are going to be big differencess between the two models. As we use data from a randomized experiment we can use a model with only a financial dummy variable included due to the nature of randomization. If an RCT was successfully conducted, the treatment assignment is statistically independent of potential outcomes, i.e.  $(Y_{0i}^*, Y_{1i}^*) \perp D_i$ . However, controlling for other regressors reduces standard errors around treatment effect. Besides, if the estimated size of the treatment effect differs with and without including covariates, then one should about the additional covariates or the initial randomization process.

Columns (1) and (2) of Table 2 show the results of the linear probability model while columns (3) and (4) yield the results marginal effects of a logit model. The results indicate that there are no statistically significant effects of a financial rewards on the probability to drop our regardless of the amount of this reward. Including control variables and using a logit model instead of a linear probability model do not change the results. The results are robust to different specifications and models used.

Table 2:

	<i>Dependent variable:</i>			
	dropout			
	<i>OLS</i>		<i>binary model</i> <i>(marginal effect)</i>	
	(1)	(2)	(3)	(4)
categoryHigh-reward	−0.053 (0.075)	−0.053 (0.073)	−0.052 (0.073)	−0.053 (0.076)
categoryLow-reward	−0.057 (0.075)	−0.069 (0.073)	−0.056 (0.073)	−0.070 (0.075)
math		−0.075*** (0.023)		−0.076*** (0.024)
fyeduc		0.007 (0.009)		0.006 (0.009)
p0		−0.328*** (0.121)		−0.322*** (0.120)
Constant	0.402*** (0.053)	0.903*** (0.163)	−0.091* (0.050)	0.418** (0.169)
Observations	249	245	249	245
Adjusted R <sup>2</sup>	−0.005	0.068		

*Note:*

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Columns (1) and (2) of Table ?? show the effects of a financial reward on the number of credits a student earned during the first year while columns (3) and (4) show the results after three years. The effect of a financial reward is zero and not statistically significant either for both outcomes.

V)

To estimate the minimum detectable effect size of the financial reward on the probability to complete all first-year courses, we are using the following formula:

$$MDE = (t_{1-\alpha/2} - t_{1-q}) \sqrt{\frac{1}{p(1-p)}} \sqrt{\frac{\sigma^2}{n}}$$

where *alpha* - significance level; *q* - power; *p* - a share of treated students;  $\sigma^2$  - variance of OLS residuals. We use conventional values for *alpha* = 0.05 and *q* = 0.8.

VI)

To increase the pass rate by 10%, we need to calculate the sample size according to this formula:

$$n = \left( \frac{t_{1-\alpha/2} - t_{1-q}}{MDE} \right)^2 \frac{\sigma^2}{p(1-p)}$$