Econometrics II. Assignment 3: Potential outcomes model, randomized experiments, and power analysis

Due date: Sunday, January 23, 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. Each subquestion has an equal number of points.

Question 1: Treatment effects

In a population, each individual is uniquely characterized by its color. There are three types of individuals, purple, blue and green individuals. A researcher wants to know how a treatment affects outcomes. The table below shows for each color how many individuals are in the treatment and control group and what the average outcomes are for individuals in the treatment and control group.

Color	Number of individual		Average outcome	
	Treated	Control	Treated	Control
Purple	100	100	9	7
Blue	75	25	13	8
Green	25	75	10	9

(i) Compute for each of the colors the treatment effect.

(ii) Compute the average treatment effect in the full population.

(iii) Compute the average treatment effect on the treated.

(iv) Give an example where the average treatment effect on the treated would be more useful to consider than the average treatment effect, and explain why.

Question 2: Lying pupils

A researcher is interested in studying how often boys and girls in high school lie. Asking about lying in a survey is not a good idea since those individuals who frequently lie may also lie when answering the survey. Therefore, the researcher develops the following research design. First, the researcher asks a pupil to memorize a number from one to six and not to tell this to the researcher. Next, the researcher asks the pupil to throw a dice and to tell if the pupil has thrown the same number as she just memorized. If pupils are honest, then $1/6^{\text{th}}$ of the pupils in a population should throw the same number as they memorized. If a fraction $x > 1/6$ of the pupils state to have thrown the correct number, then the fraction of pupils that lie is $\frac{x-1/6}{1-1/6}$.

The researcher decides to make three groups. The first treatment group gets fruit if they state that they have thrown the correct number. The second treatment group gets a candy bar if they state to have thrown the correct number. And third the control group gets nothing for throwing the correct number.

The file `assignment3.csv(\.dta)` contains 203 pupils who have participated in the experiment. It includes the following variables:

Variable	Description
id	id
gender	gender
grade	grade
important	finding school or sports more important
better	being better in math or languages
preferredhand	being left or right-handed
siblings	1 if a person has (older) siblings
youngestchild	1 if a person is the youngest sibling
oftenexpelled	1 if a person is expelled often from a lecture
correct	1 if a person has the correct number on the dice
treatment	treatment group

Correct is the outcome variable and *Treatment* indicates the treatment groups.

(i) Compute the fraction of pupils in all three groups (control, fruit, and candy bar) that have the number correct and that are expected to lie. Show within a table whether pupil characteristics are balanced over the treatment groups.

(ii) Use the linear probability model to regress the dummy variable for having the correct number on the dice on the assignment to the three treatment groups. Do you detect significant lying? Show how robust you results are to including additional covariates.

(iii) The researcher is interested in testing if boys and girls behave differently. Write down and estimate a model that can test if boys lie more than girls and whether boys and girls respond differently to incentives (e.g. treatment).

(iv) Estimate your preferred model specification (including pupil characteristics if this improves your model) and write down your main conclusions.

Prior to conducting the experiment, the researcher performed some power calculations.

(v) Given the sample size and the estimates you have obtained above, what would be the minimum detectable effect size of this experiment?

(vi) Initially, the researcher expected that no pupil would lie if there is nothing at stake (control group) and that a quarter of the student lie for a candy bar. How large should the sample size of the experiment have been in that case?

(vii) Students assigned to the control group provide counterfactuals to both the fruit and the candy bar treatment. Show that by assigning more students to the control group than to each treatment group the same MDE can be achieved with fewer students in the experiment.