Assignment 3

David Gyaraki, Thao Le

Contents

```
#{r setup q1, echo=TRUE, results='hide'} ## load packages #if(!require(pacman)){install.pack
# #p_load(devtools,tidyverse,dplyr,ggplot2,latex2exp, #        sampleSelection,
quantreg, plm, nlme) # ##load data #dfData = read.csv("assignment2a_2023.csv")
#attach(dfData) #
```

1 Question 1

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

1.1 (i)

The treatment effect in theory is the difference between the outcomes if the individual is treated versus if the individual is not treated. Suppose that for individual i, the treatment effect is defined as:

$$TE_i = \Delta_i = y_{i,d=1} - y_{i,d=0}, \tag{1}$$

where the y marks the outcome and d is the dummy whether individual i was treated $(d = 1)$ or not $(d = 0)$. However, this is rarely possible to be observed for each individual, as treatment is generally considered mutually exclusive, so an individual is either treated or not. Therefore, in this example we cannot calculate TE measures for individuals within color groups, but we can pretend as if one color corresponds to one observation, and calculate the three treatment effects across the color groups.

Then, the treatment effect for "observation" purple is $9 - 7 = 2$, the treatment effect for "observation" blue is $13 - 8 = 5$ and the treatment effect for "observation" green is $10 - 9 = 1$.

1.2 (ii)

Using the assumption of Section 1.1, we can infer here that we are looking for the average treatment effect of the full population including all colors. The average treatment effect is generally given by:

$$ATE = E[\Delta] = E[y_{d=1} - y_{d=0}] = E[y_{d=1}] - E[y_{d=0}] = \frac{1}{N_T}\Sigma_{i=1}^{N_T} y_{i,d=1} - \frac{1}{N_{NT}}\Sigma_{i=1}^{N_{NT}} y_{i,d=0},$$

(2)

where N_T is the number of individuals in the treatment group while N_{NT} is the number of individuals in the control group. Our data in the exercise does not contain information again on the individual outcomes, however we can use the average outcomes of treatment and control in each color group to calculate the total $E[y_{d=1}]$ and $E[y_{d=0}]$ on the full population. Therefore, the ATE for the whole population can be calculated as:

```
E_treatment <- (9*100 + 13*75 + 10*25) / (100+75+25)
E_control <- (7*100 + 8*25 + 9*75) / (100+75+25)

ATE = E_treatment - E_control
ATE
```

[1] 2.75

1.3 (iii)

1.4 (iv)

The ATE measure generally describes the expected gain in y achieved by treating a random member i from the population, i.e. how much one benefits from being selected for the treatment compared to people who were not. On the other hand, ATET describes the average gain achieved by the treatment for the treated group, i.e. the comparison is not to the population but peer-to-peer, what is the expected benefit for those who are selected. The ATET is more helpful if we are not mainly interested in the potentially positive effect of the treatment for those who are treated versus those who are not, but rather the magnitude of these positive effects. Suppose that we have an experiment where the government announces a new plan to introduce an additional level of health insurance, where the own risk cost would be cut in half, in order to investigate the effects of these on household savings. Arguably, cutting the own risk cost in half without changing the insurance monthly premiums would most likely have a positive effect on the wealth for those who are involved in the initial study, but if the government is rather interested in measuring the average savings surplus this would create for households, we would be more interested in the ATET measure as this policy will ideally be introduced for everyone later on and we are solely interested in the average savings surplus this would create for everyone involved.

2 Question 2

3

```
#{r load data q3, echo=TRUE, results='hide'} #dfData2 = read.csv("assignment2b_2023.csv")
#attach(dfData2) # #dfData2 <- na.omit(dfData2) #
```

3.1 (i)

3.2 (ii)

3.3 (iii)

3.4 (iv)

3.5 (v)

3.6 (vi)

3.7 (vii)