REGRESSION ANALYSIS AND CAUSALITY WITH R

Treatment Effects

João Cerejeira[1]   Miguel Portela[1,2,3]

[1]NIPE – UMinho
[2]IZA, Bonn
[3]Banco de Portugal

October 26, 2021

Introduction: selection bias

Data exhibit a selection bias, because some people chose (or self-selected) to go to the hospital and the others did not.
When membership in the treated group is in part determined by choice, then the sample is not a random sample.

We would like to randomly assign items to a treatment group, with others being treated as a control group.
We could then compare the two groups.
The key is a randomized controlled experiment.

The difference estimator

Define the indicator variable d as:

$$d_i = \begin{cases} 1 \text{ individual in treatment group} \\ 0 \text{ individual in control group} \end{cases}$$

The model is then:

$$y_i = \beta_1 + \beta_2 d_i + u_i \tag{1}$$

And the regression functions are:

$$E(y_i) = \begin{cases} \beta_1 + \beta_2 \text{ if } i \text{ in treatment group, } d_i = 1 \\ \beta_1 \text{ if } i \text{ in control, } d_i = 0 \end{cases}$$

The OLS estimator for β_2 is:

$$b_2 = \frac{\sum_{i=1}^{N}\left(d_i - \overline{d}\right)\left(y_i - \overline{y}\right)}{\sum_{i=1}^{N}\left(d_i - \overline{d}\right)^2} = \overline{y}_1 - \overline{y}_0$$

The estimator b_2 is called the difference estimator, because it is the difference between the sample means of the treatment and control groups.

The difference estimator can be rewritten as:

$$b_2 = \beta_2 + \frac{\sum_{i=1}^{N} \left(d_i - \overline{d}\right)\left(e_i - \overline{e}\right)}{\sum_{i=1}^{N} \left(d_i - \overline{d}\right)^2} = \beta_2 + \left(\overline{e}_1 - \overline{e}_0\right)$$

To be unbiased, we must have:

$$E\left(\overline{e}_1 - \overline{e}_0\right) = E\left(\overline{e}_1\right) - E(\overline{e}_0) = 0$$

If we allow individuals to "self-select" into treatment and control groups, then $E\left(\overline{e}_1\right) - E(\overline{e}_0)$ is the selection bias in the estimation of the treatment effect.

We can eliminate the self-selection bias if we randomly assign individuals to treatment and control groups, so that there are no systematic differences between the groups, except for the treatment itself.