Swiss Institute of Artificial Intelligence   **Minwook Ahn**
MBA AI/BigData   **Problem Set 3**
[STA501] Data-based Decision Making   **September 30, 2021**

Question 1

[Answer] To see that X causes Y, other effects besides X does not causes Y (ceteris paribus). For example, in the statement of "He was hired for the job because he is a man. A woman would be less likely to get the job", X is gender and Y is getting the job. From a lot of data randomly selected from population, the factors besides gender should have no influence to Y on average. In order to estimate its causal effect, we have to get ATE_1:

$$ATE_1 = E(y_{i1}|d_i = 1) - E(y_{i0}|d_i = 1) \tag{1}$$

This is the causal effect of gender on those who get the job. However, the observable quantity in reality is like the following:

$$E(y_{i1}|d_i = 1) - E(y_{i0}|d_i = 0) \tag{2}$$

By modification, we have:

$$E(y_{i1}|d_i = 1) - E(y_{i0}|d_i = 0) = E(y_{i1}|d_i = 1) - E(y_{i0|}d_i = 1) + E(y_{i0}|d_i = 1) - E(y_{i0}|d_i = 0) \tag{3}$$

where we have the following selection bias:

$$E(y_{i0}|d_i = 1) - E(y_{i0}|d_i = 0) \tag{4}$$

Question 2

[Answer] False. If W is randomly assigned for Y, then it can be a valid instrument for T. However, W is randonly assigned to treatement T.

Question 3

[Answer] False. There are two reasons. First, the doctor should consider placebo effect. Second, the regression assumes the time effect is the same for the treated and non-treated. We also have to consider time effect in order not to wrongly estimate ATE.

Question 4-1

[Answer] The simple version of linear regression will be:

$$Y = \beta_0 + \beta_1 x_1 + \epsilon \tag{5}$$

where x_1 is a dummy variable to describe the market access (1: restricted or close to border, 0: not restricted). However, there are other factors to be considered such as substitution of city product (x_2), the amount of tradable product(x_2), transport cost of product to other cities (x_3), and size of city (city population, x_4).

Question 4-2

[Answer] By considering the factors described above, the equation will be:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \lambda x_5 + \epsilon \tag{6}$$

where the others are the same as described in 4-1 and x_5 describes the common trend of economy in time. In this case, β_1 can show us how much the economical growth of a city is influenced by market access to that city. In other word, β_1 will be ATE.

Question 4-3

[Answer] If we have more observations on y both before and after the construction of the Iron Curtain, it is helpful to consider the time effect, which makes us under-estimate ATE of market access on economy.

Question 5

Question 6-1

[Answer] γ_3 of non-white and γ_4 of male show the opposite ATE of γ_3 of white and γ_4 of female. Especially even though their values ($\gamma's$) are the same but their signs will be opposite.

Swiss Institute of Artificial Intelligence
MBA AI/BigData
[STA501] Data-based Decision Making

Minwook Ahn
Problem Set 3
September 30, 2021

Question 6-2

[Answer] The interaction variable Z can remove the correlative effect between age and white on earnings. Therefore, we can induce the independent effect of age or white on earnings. The same as age and female.

Question 6-3

[Answer] If there is a variable with heteroskdecity, then the regressor and disturbance may be statically related. I suspect tenure.

Question 7-1

[Answer]

$$E(Y_i|T_i = 1) - E(Y_i|T_i = 0) = E(Y_{i1}|T_i = 1) - E(Y_{i0}|T_i = 0) = E(Y_{i1}|T_i = 1) - E(Y_{i0}|T_i = 1) + E(Y_{i0}|T_i = 1) - E(Y_{i0}|T_i = 0) \tag{7}$$

where $E(Y_{i0}|T_i = 1) - E(Y_{i0}|T_i = 0)$ is selection bias.

Question 7-2

[Answer]

Question 7-3

[Answer]

$$ATE = E(Y_{it,1}|T_i = 2) - E(Y_{it-1,0}|T_i = 1) = \beta_0 + \beta_1 + \beta_2 + \beta_3 - \beta_0 = \beta_1 + \beta_2 + \beta_3 \tag{8}$$

Question 7-4

[Answer] I think that β_5 can help to reduce time effect.