# Basics of Causal Inference

Qiguang Yan & Kenny Zhang (Mentor)

As part of the UW Statistics and Probability Association's Directed Reading Program, I explored the development and notable methods of causal inferences with my mentor Kenny Zhang. We read *The book of why: The new science of cause and effect* by Judea Pearl. Humans are causal beings, and we view the world in terms of causal relationships. Pearl believes that the key to creating human like artificial intelligence is understanding and develop tools to derive and analyze causal relationships.

A confounding variable induces changes in both the explanatory variable and response. To measure the causal effect of the explanatory variable on the response variable, we must control for the effects of confounders. Causal relationship can be established from randomized control trials. Assuming we want to test the effect of a fertilizer on crop yields. If we split the field and apply fertilizer to one side, we introduce soil fertility as a confounder. If we apply fertilizer on the field this year and use the yield from last year as control, we introduce climate as a confounder. If we assign the fertilizer randomly, the only variable that induces changes to the explanatory variable is the random number itself. Thus, the causal relationships between confounders and the explanatory variable are blocked. We can then estimate the average causal effect free of the influence from the confounders.

Randomized control trials have several drawbacks. Intervention can sometimes be unethical or even impossible. It is difficult to recruit subjects who are not volunteers for experiments. Additionally, randomized control trials can be very expensive. Thus, establishing causal relationship from observational data alone is desirable. Pearl discussed three methods that could accomplish this. First, we must understand the three types of junctions in causal diagrams.

Assuming fire creates smoke and smoke triggers alarm, smoke is a mediator between fire and alarm. If smoke in a room is fixed, fire and alarm are independent. The state of fire does not provide information about the state of the alarm, vice versa.

Mediator:

Fire → Smoke → Alarm

Assuming older children have larger shoe sizes and higher reading ability, age is a confounder between shoe sizes and reading ability. For children at the same age, shoe sizes and reading ability are independent. The shoe size of a child does not provide information about the reading ability of that child, vice versa. Giving a child larger shoe does not improve reading ability.

Confounder:

Shoe sizes ← Age → Reading ability

Assuming talent and beauty both contribute to success in acting professions. Success is a collider between talent and beauty. Talent and beauty are independent. However, at a fixed level of success, people with higher talent tend to have lower level of beauty, people with higher level of beauty tend to have lower level of talent.

Collider:

Talent → Success ← Beauty

Back-door adjustment attempts to control cofounders directly or block their effects by controlling their mediators. A back door path has arrow pointing both towards the explanatory and response variable. If there is a collider in the back door path, it is already blocked. If there are only the confounder and its mediators in the back door path, controlling any variable in the path will achieve the desirable outcome. Controlling colliders should be avoided because controlling colliders opens the back door path. Controlling a variable means we calculate the causal effect at every value of the confounder. An example would be calculating the causal effect between shoe sizes and reading ability for children of a specific age, then move on to the next specific age.

Sometimes, confounders are unobserved, such as hypothetical genes, motivation, lifestyle, etc. We cannot use the back door adjustment and control for these confounders. Front-door adjustment could be useful in some of these situations. Front-door adjust assumes that there is a mediator between the explanatory variable and the dependent variable, and there is no causal relationship between the mediator and confounder. We can measure the causal effect between explanatory variable and the mediator because there are no confounders between them. We can also calculate the causal effect between the mediator and the dependent variable because there are no confounders between them. We can then estimate the causal effect between the explanatory variable and response.

Front-Door Adjustment:

Unobserved Confounder

↓ ↓

Explanatory → Mediator → Response

Instrumental variables are another useful method. An instrument is a variable that induces changes in the explanatory variable while having no direct causal relationship with confounders or the response variable. The instrument can cause changes in the response variable only through the explanatory variable as the mediator. We can calculate the causal effect of the instrument on the explanatory variable, and the instrument on the response variable. We can then estimate the average causal effect of the explanatory variable on the response variable.

Instrumental Variables:

Unobserved Confounder

↓ ↓

Instrument → Explanatory → Response

Reference:

Pearl, J., & Mackenzie, D. (2020). *The book of why: The new science of cause and effect*. Hachette Books Group.