

Short summaries of what I read so far.

Jaemin Oh

November 20, 2022

1 Causal Inference

1.1 Holland, 1986 [1]

The very basic of Rubin's model was explained, and the distinction between associational inference and causal inference was provided.

Let U be the population and $u \in U$ be a unit. Three variables Y , W , and C are given, where Y is the variable to be analyzed (response variable), W is a (potential) cause of Y , and C is an attribute. They are functions from U to \mathbb{R} , and their distributions are given by the relative frequency on U . As a function, W and C are the same. However, they are different in the sense that we cannot do randomized experiment with C and can do with W . For example, a clinical surgery can be W and the gender can be C . This property distinguishes the cause from the attribute.

In this setting, the associational inference is focused on $E(Y|W)$ or $E(Y|C)$. In other words, the discovery of the way that Y is related to W or C will be satisfactory. On the other hand, in causal inference, a direct comparison between treatment and control for each unit is required. This cannot be done in practice, since any unit cannot receive both treatment simultaneously. But we can do counterfactual imaginations that lead additional functions $\{Y_w\}_{w \in I}$ which is called potential outcomes. To overcome practical issue and estimate the causal effect, the researcher should design the study to approximate randomized experiment, which is the simplest setting. Note that, in randomized experiment, $Y_w =_d Y|W = w$ by consistency and ignorability (missing at random).

References

- [1] Paul W. Holland. Statistics and causal inference. *Journal of the American Statistical Association*, 81(396):945–960, 1986.