

Bob Bell

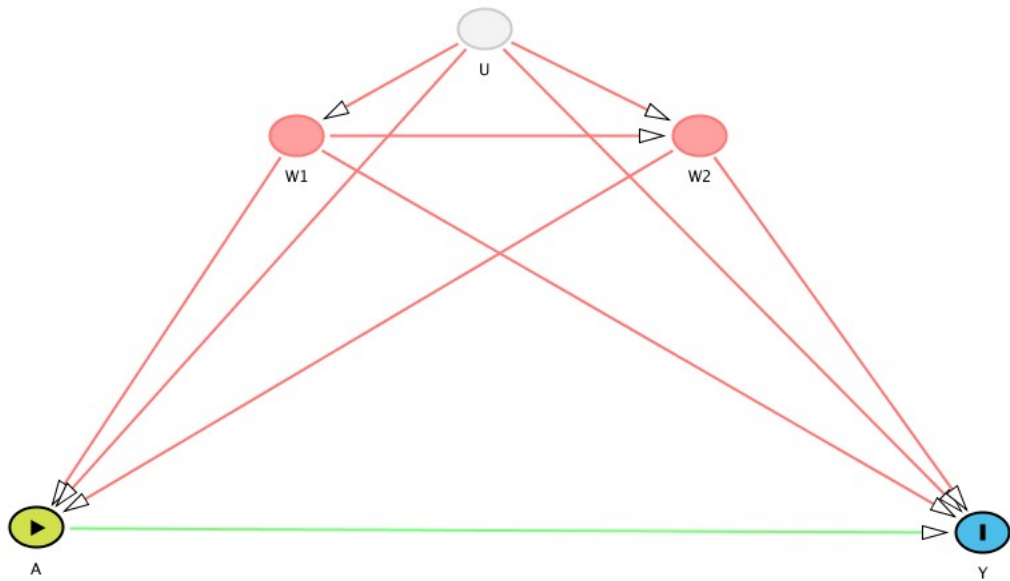
10.21.2013

PH252D

R Assignment 1

1 Background

1. Directed Acyclic Graph:



2. There are no exclusion restrictions because the structural equation for every endogenous variable includes in its parent set all endogenous variables before it (with respect to time ordering).

3. There are no independence assumptions, as we have not assumed any exogenous/background (U) variables are independent from other U's.

4. Counterfactual outcome of interest are $(Y_a : a \in \mathcal{A} = \{0, 1\})$ where $Y_a = f_Y(W, a, U_Y)$. Y_0 is the counterfactual child's weight gain in pounds at the study termination if the child had received the standard supplement. Y_1 is the counterfactual child's weight gain in pounds at the study termination if the child received RUTF.

5. The target causal parameter is the average causal effect given by the following notation:

$$\Psi^F(P_{U,X}) = E_{U,X}(Y_1) - E_{U,X}(Y_0) = E_{U,X}[f_Y(W1, W2, 1, U_Y)] - E_{U,X}[f_Y(W1, W2, 0, U_Y)]$$

Average treatment effect is the difference in the expected counterfactual weight gain of all children at the end of the study if all children received RUTF and the expected value of the counterfactual weight gain of all children at the end of the study if all children received the standard supplement.

6. Observed data is generated by the data-generating system contained by our SCM. Therefore, $O \subseteq X$. The distribution of the background variables P_U and the structural equations F identify the distribution of X and thus the distribution for O . Our observed data was generated by sampling n i.i.d. times from the data generating system specified in our causal model, giving us O_1, O_2, \dots, O_n from the probability distribution P_0 . The statistical model has no exclusion restrictions or independence assumptions. Thus, it is non-parametric. This true, unknown distribution of the observed data and the statistical model can be translated into the structural causal model \mathcal{M}^F .

7. Using the backdoor criterion, the target causal parameter is not identified. I need to make some assumptions to establish identifiability. If $U_A \perp U_Y$, then we can establish identifiability controlling for $\{W_1, W_2\}$. This original SCM with additional assumptions is denoted \mathcal{M}^{F*} .

8. Target parameter (statistical estimand) is:

$$\begin{aligned} \Psi(P_0) &= E_0[E_0(Y | A = 1, W_1, W_2) - E_0(Y | A = 0, W_1, W_2)] \\ &= \sum_{w_1, w_2} [E_0(Y | A = 1, W_1 = w_1, W_2 = w_2) - E_0(Y | A = 0, W_1 = w_1, W_2 = w_2)] P_0(W_1 = w_1, W_2 = w_2) \end{aligned}$$

9. Positivity Assumption:

$$\min_{a \in A} P_0(A = a | W_1 = w_1, W_2 = w_2) > 0 \text{ (i.e.) } 0 < P_0(A = 1 | W_1 = w_1, W_2 = w_2) < 1$$

for all w_1, w_2 for which $P_0(W_1 = w_1, W_2 = w_2) > 0$. I think the positivity assumption is reasonable here. Theoretically, for all covariate pairs, it is possible to observe children receiving both the RUTF and standard supplements. Practically, I could also make this assumption as we only have 4 covariate pairs.

2 A specific data generating process

1. Assemble components of target causal parameter and estimate:

$$P_{U,X}(W_1 = 1) = 0.20$$

$$P_{U,X}(W_1 = 0) = 0.80$$

$$E_{U,X}(W_2 | W_1 = 1) = \text{expit}(0.5 * 1) = \frac{1}{1 + e^{-0.5}} \approx 0.62246$$

$$E_{U,X}(W_2 | W_1 = 0) = \text{expit}(0.5 * 0) = \frac{1}{1 + e^0} \approx 0.50000$$

$$E_{U,X}(W_2) = 0.62246 * 0.20 + 0.5 * 0.8 \approx 0.52449$$

$$\begin{aligned} E_{U,X}(Y_1) - E_{U,X}(Y_0) &= E_{U,X}[4 * 1] + E_{U,X}[0.7 * W_1] - E_{U,X}[2 * 1 * W_2] + \\ E_{U,X}[U_Y] - E_{U,X}[4 * 0] - E_{U,X}[0.7 * W_1] + E_{U,X}[2 * 0 * W_2] - E_{U,X}[U_Y] \\ &= E_{U,X}[4 * 1] - E_{U,X}[2 * 1 * W_2] \\ &= 4 - 2 * E_{U,X}[W_2] \end{aligned}$$

$$= 4 - 2 * 0.52449$$

$$= 2.9510$$

2. Interpret this estimand: On average, for all children, the impact of the RUTF increases the child's weight gain at the end of the study by 2.9510 pounds.